UNIVERSITY
OF OSLO

Nils-Ole Stutzer

Towards the first line intensity map of cosmic carbon monoxide with the COMAP Experiment

Thesis submitted for the degree of Philosophiae Doctor

Institute of Theoretical Astrophysics
The Faculty of Mathematics and Natural Sciences

University of Oslo

2025

„Die Tatsache, dass man etwas über den Himmel weiß, ändert an seinem Zauber nichts."

Harald Lesch

Summary

Over the last century, cosmology has developed into a high-precision science. For instance, cosmic microwave background (CMB) and galaxy surveys have constrained the physics of the Universe at early and late times, respectively, in great detail. However, periods like the Epoch of Reionization and the Epoch of Galaxy Assembly remain data-starved, and details about the physics governing these epochs are highly uncertain. To fill this gap, the technique of line intensity mapping (LIM) collects all light from bright and faint sources of some redshifted emission line without resolving any individual sources, resulting in an unbiased 3D map of the large-scale structure in the Universe up to high redshifts. The *Carbon monOxide Mapping Array Project (COMAP)*, the main subject of this thesis, is currently in its Pathfinder stage and is one of the leading LIM experiments. It is aiming to map the Epoch of Galaxy Assembly (cosmic redshift z = 2–3) and Epoch of Reionization (z = 6–8), using the rotational emission lines of carbon monoxide (CO(1–0) and CO(2–1)), measured at 26–34 GHz.

This thesis contains work on two successive generations of the COMAP data analysis pipeline, called COMAP Early Science (ES) and Season 2 (S2). The work entails aspects of both the low-level analysis of the raw time- and frequency-ordered data and the high-level pipeline that computes power spectra to infer the physics around cosmic noon.

In the first generation of the pipeline, developed for COMAP ES, we filter out systematic effects from the data, such as continuum foreground, $1/f$ gain fluctuations, ground pickup, and standing waves in the telescope optics. Furthermore, we calibrate the data and bin the time- and frequency-ordered data into maps. Subsequently, we compute cross-power spectra between maps of different feeds and elevations, each containing independent detector- and elevation-specific systematic errors, to construct a robust and sensitive estimator for the extragalactic CO(1–0) power spectrum. The resulting feed-feed pseudo-cross-power spectra (FPXS), containing about a year's worth of data, are the first direct constraints on the CO(1–0) clustering power spectrum found in the literature.

In the second-generation pipeline, we build on the lessons learned from COMAP ES and introduce several new filters to mitigate new systematic effects uncovered by integrating the increased data volume of S2. Specifically, we find two new highly pointing- and frequency-correlated systematic effects, coined the turn-around (TA) and start-of-scan (SoS) effects, which likely originate in standing waves in the optical path or electronics of the telescope. To address these new systematic effects, we develop two new principal component analysis (PCA) filters, one acting on the time-domain in each detector and the

other in the map-domain. After filtering, the maps are consistent with noise expectations, and the noise integrates down with additional data.

To increase the robustness against the TA and SoS, which seem to correlate to specific groups of feeds, we modify the S2 power spectrum methodology only to compute feed-group pseudo-cross-power spectra (FGPXS). We improve the accuracy of our power spectrum uncertainties by estimating the S2 power spectrum errors from the power spectra of randomized half-difference maps that inherit all noise properties and biases due to the low-level pipeline filters from the data. Combining this with a set of 312 difference-map null tests and new and improved transfer function estimators, we ensure that the final data product of COMAP S2 is consistent with instrumental noise expectations and is corrected for known biases. The COMAP S2 power spectrum provides the world's tightest direct constraints on the 3D CO(1–0) clustering power spectrum at $z = 2$–3 at the time of writing. The upper limits on the CO(1–0) power spectrum at 95 % confidence are an order of magnitude deeper than those of COMAP ES and the CO Power Spectrum Survey (COPSS; the only comparable CO(1–0) LIM survey with published results). As such, we find that COMAP is en route to secure the world's first cosmic CO detection at the Epoch of Galaxy Assembly.

Next, we consider some algorithmic improvements, developed primarily for the BeyondPlanck and Cosmoglobe CMB projects, that can be adapted in a future high-precision COMAP LIM pipeline. That is, an iterative Bayesian approach is a more well-motivated framework than the classic COMAP pipeline, as it can directly map out the posterior probability space of all experimental parameters and reveal non-trivial correlations between parameters. We show the power of the `Commander3` framework by using a set of simulations of the Planck LFI 30 GHz data to validate the `Commander3` global Bayesian Gibbs sampler, including gain and correlated noise estimation as well as mapmaking. The framework performs well and recovers all input parameters as expected within the experimental uncertainties. It also provides a detailed overview of all non-trivial uncertainties and correlations of the system. Incorporating COMAP into the `Commander` framework will be an important future goal.

Finally, in the last piece of work of this thesis, we propose a maximum likelihood method that directly takes the Planck HFI bolometer transfer function into account in the mapmaking stage without having to use any deconvolution steps. This results in an effective beam ellipticity and full-width-at-half-maximum that are, respectively, 64 % and the 2.3 % decreased as compared to the deconvolution performed by the original Planck HFI team. The main contribution to this effort was to perform an independent validation of the algorithm using simplified 1D toy models of the Planck HFI 143 GHz data.

Samandrag

Gjennom det siste hundreåret har kosmologi utvikla seg til å bli ei presisjonsvitskap. Til dømes har granskingar av den kosmiske mikrobølgjebakgrunnen og kartlegging av galaksar auka kjennskapen vår om fysikken i både det tidlege og seine universet. Men periodar som reioniseringsepoken og galaksedanningstida er framleis prega av lite data, og detaljar omkring fysikken som styrer desse epokane er svært usikre. For å fylle dette hòlet nyttar teknikken spektraldjupnekartleggjing (LIM) seg av alt ljos frå både ljossterke og svake kjelder av ei bestemt raudforskuva utslippslinje utan å oppløyse individuelle kjelder. Dette skapar eit forventingsrett 3D-kart over storskalastrukturen i universet opp til høge raudforskuvingar. *The Carbon monOxide Mapping Array Project (COMAP)*, hovudtemaet i denne avhandlinga, er for tida i sin Pathfinder-fase og er eit av dei leiande LIM-eksperimenta. Målet er å kartleggje galaksedanningstida (kosmisk raudforskuving z = 2–3) og reioniseringsepoken (z = 6–8) gjennom bruk av rotasjonsutslippslinjene av karbonmonoksid (CO(1–0) og CO(2–1)) målt ved 26–34 GHz.

Denne avhandlinga inneheld arbeid på to etterfølgande generasjonar av COMAP databehandlingssystemet, kalla COMAP Early Science (ES) og Sesong 2 (S2). Det omfattar lågnivåanalysen av dei rå tid- og frekvensordna dataa og høgnivåanalysen som reknar ut effektspektra for å avgjere lovane til fysikken rundt tida av høgaste stjerneskaping.

I den fyrste generasjonen av databehandlinga, utvikla for COMAP ES, filtrerer me bort systematiske effektar frå dataa, slik som kontinuumsframgrunn, $1/f$-forsterkingsfluktuasjonar, jordplukk, og ståande bølgjer i teleskopoptikken. Deretter kalibrerer me dataa, og projiserer dei frekvens- og tidordna dataa til kart på himmelen. So reknar me ut krysseffektspektra mellom kart av ulike mottakarar og pekehøgder, sidan dei kvar inneheld uavhengige detektor- og høgdespesifikke systematiske feil. Med dette lagar me ein robust og kjenslevar estimator for det utanomgalaktiske CO(1–0) effektspekteret. Dei resulterande mottakar-mottakar pseudokrysseffektspektra (FPXS), som inneheld om lag eit års data, er dei fyrste direkte avgrensingane på CO(1–0) opphopingsspekteret i forskingslitteraturen.

I den andre generasjonen av databehandlinga byggjer me vidare på erfaringane frå COMAP ES og innfører fleire nye filter for å dempe nye systematiske effektar som kjem fram når me legg saman den auka datamengda frå S2. Spesielt finn me to nye svært pekings- og frekvenskorrelerte systematiske effektar, kalla vendingseffekten (TA) og sveipbyrjingseffekten (SoS). Sannsynlegvis stammar dei frå ståande bølgjer i optikken eller elektronikken til teleskopet. For å handtere desse nye systematiske effektane, utviklar me to

nye hovudkomponentanalyse filter (PCA-filter), eitt som verkar på tidsdomenet i kvar detektor og det andre i kartdomenet. Etter filtrering er karta konsistente med støyforventingar, og støyen integrerer ned med meir data.

For å gjere oss meir motstandsdyktig mot TA- og SoS-effektane, som ser ut til å korrelere med bestemte grupper av mottakarar, modifiserer me metodologien for S2 effektspekter til å berre rekne ut mottakar-gruppe pseudokrysseffektspektra (FGPXS). For å auke grannsemda av effektspekterutryggleikane, estimerer me feila i S2 effektspekter ved hjelp av effektspekter av tilfeldige halv-differansekart, som arvar alle dei rette støyeigenskapane og forventingsskeivskapane frå lågnivåanalysefiltrene. Ved å kombinere dette med eit sett av 312 differansekartnulltestar, so vel som nye og betre estimatorar av overføringsfunksjonane, sikrar me at dataproduktet frå S2 er konsistent med instrumentale støyforventingar og er korrigert for kjende skeivskapar. Det resulterande COMAP S2 effektspekteret gjev strammaste direkte avgrensingar på 3D CO(1–0) opphopingsspekteret i verda. Me får ein storleiksorden djupare øvre grenser på CO(1–0) effektspekteret ved 95 % tryggleik enn dei frå COMAP ES og CO Power Spectrum Survey (COPSS; den einaste samanliknbare CO(1–0) LIM-granskinga med publiserte resultat). Dermed finn me at COMAP er på veg til å sikre den fyrste kosmiske CO-målinga i verda frå galaksedanningstida.

Vidare vurderer me nokre algoritmeforbetringar, utvikla for BeyondPlanck og Cosmoglobe CMB-prosjekta, som kan tilpassast ei framtidig høgpresisjons COMAP LIM-analyse. Det vil seie, ei iterativ gjennomgåande Bayesiansk analyse er ein betre grunngitt framgangsmåte enn den klassiske COMAP analysen. Den kartleggjer direkte a-posteriori-sannsynfordelinga til eksperimentelle parametrar og avdekkjer ikkje-openberre korrelasjonar mellom parametrar. Me viser styrken til `Commander3`-koden ved å bruke eit sett med simuleringar av Planck LFI 30 GHz data for å validere `Commander3` Gibbs-prøvetakinga, inkludert forsterkings- og korrelert støyestimering i tillegg til kartlaginga. Koden fungerer godt og attskapar alle byrjingsverdiar som forventa innan eksperimentelle utryggleikar. Den gjev i tillegg oversikt over alle ikkje- openberre utryggleikar og korrelasjonar i systemet. Å innleme COMAP i `Commander3`-koden vil vere eit viktig framtidig mål.

Til slutt, i det siste bidraget til denne avhandlinga føreslår me å handtere Planck HFI bolometer-overføringsfunksjonen direkte i ein optimal sannsynmaksimeringsmetode. Dette gjev ein effektiv stråleelliptisiteten og full-bredde-ved-halv-maksimum som er høvevis 64 % og 2.3 % mindre i høve til den opphavelege Planck HFI metoden. Hovudbidraget til dette arbeidet er å utføre ei uavhengig validering av algoritmen ved hjelp av forenkla 1D modellar av Planck HFI 143 GHz dataa.

Acknowledgments

This work is supported by the Bits2cosmology and Cosmoglobe ERC consolidator grants 772253 and 819478, and the Institute of Theoretical Astrophysics at the University of Oslo. This material is based upon work supported by the National Science Foundation under Grant Nos. 1517108, 1517288, 1517598, 1518282, 1910999, and 2206834, as well as by the Keck Institute for Space Studies under "The First Billion Years: A Technical Development Program for Spectral Line Observations".

I want to thank my supervisors, Hans Kristian and Ingunn, for all their advice and support during the last couple of years and for allowing me to be part of the group/family. I also want to thank my co-supervisor, Håvard, for answering all my stupid questions, advising me during my Ph.D. and M.Sc. before that, and teaching me how to play D&D properly.

Thanks to my colleagues and friends in the COMAP collaboration. It is an honor to be part of COMAP! May the future shine bright in redshifted CO line emission.

Thank you to my office mates, colleagues, and friends Jonas and Sigurd for all the good discussions, collaboration, and help. I hope you had as much fun as I while discussing the fundamental mysteries of this world while biking and hiking through the endless woods of Oslo.

During my time in the CMB&CO group, I have had the best colleagues and friends imaginable. All of you, Ana, Jonas, Sigurd, Metin, Håvard, Marie, Duncan, Eirik, Mathew, Raelyn, Daniel, Maksym, Artem, and Unni, are great colleagues and even better friends. The same goes for my other colleagues and friends at ITA and UiO; you are all truly awesome. Thanks to Ana, Bendix, Louisa, Metin, Eilif, and Raelyn for being the best climbing crew in the local Universe.

Danke an meine Eltern Alexa und Jens für eure Unterstützung und Ermutigung meinen Weg zu finden, auch wenn dieser zeitweise steil erschien. Ich verspreche euch, ihr müsst nie wieder im Flugzeug aufstehen, wenn nach einem Doktor an Bord gefragt wird: Ich mache das schon! Einen besonderen Dank an dich Joachim, dafür dass du mir in jungen Jahren mit deiner Sternwarte gezeigt hast wie wunderschön das Universum ist. Danke an meine Brüder Lars, Lennart und Adrian, und meinen Kater Ramses, dafür dass ich immer auf euch zählen kann. Ihr seid alle Klasse!

To all of you, thank you for making me the person I am today,

Nils-Ole Stutzer

Oslo, March 2025

List of papers

Paper I

Foss, M. K., et al. 2022, arXiv:2111.05929, ApJ, 933, 184, COMAP Early Science. III. CO Data Processing

Paper II

Ihle, H. T., et al. 2022, arXiv:2111.05930, ApJ, 933, 185, COMAP Early Science. IV. Power Spectrum Methodology and Results

Paper III

Lunde, J. G. S., et al. 2024, arXiv:2406.07510, A&A, 691, A335, COMAP Pathfinder – Season 2 results: I. Improved data selection and processing

Paper IV

Stutzer, N. O., et al. 2024, arXiv:2406.07511, A&A, 691, A336, COMAP Pathfinder – Season 2 results: II. Updated constraints on the CO(1–0) power spectrum

Paper V

Chung, D. T., et al. 2024, arXiv:2406.07512, A&A, 691, A337, COMAP Pathfinder – Season 2 results: III. Implications for cosmic molecular gas content at z$\sim$3

Paper VI

Brilenkov, M., et al. 2023, arXiv:2209.04437, A&A, 675, A4, BEYONDPLANCK. IV. Simulations and validation

Paper VII

Basyrov, A., Stutzer, N. O., Lunde, J. G. S., Eriksen, H. K., Gjerløw, E., Watts, D. J., & Wehus, I. K. 2024, A&A, 688, A166, Optimal bolometer transfer function deconvolution for CMB experiments through maximum likelihood mapmaking

Additional papers published, but not included in this thesis

COMAP papers

1. Cleary, K. A., et al. 2022, arXiv:2111.05927, ApJ, 933, 182, COMAP Early Science. I. Overview

2. Lamb, J. W., et al. 2022, arXiv:2111.05928, ApJ, 933, 183, COMAP Early Science. II. Pathfinder Instrument

3. Chung, D. T., et al. 2022a, arXiv:2111.05931, ApJ, 933, 186, COMAP Early Science. V. Constraints and Forecasts at z 3

4. Rennie, T. J., et al. 2022, arXiv:2111.05932, ApJ, 933, 187, COMAP Early Science. VI. A First Look at the COMAP Galactic Plane Survey

5. Breysse, P. C., et al. 2022, arXiv:2111.05933, ApJ, 933, 188, COMAP Early Science. VII. Prospects for CO Intensity Mapping at Reionization

6. Chung, D. T., et al. 2021, arXiv:2104.11171, ApJ, 923, 188, A Model of Spectral Line Broadening in Signal Forecasts for Line-intensity Mapping Experiments

7. Dunne, D. A., et al. 2024, arXiv:2304.09832, ApJ, 965, 7, COMAP Early Science. VIII. A Joint Stacking Analysis with eBOSS Quasars

8. Harper, S. E., et al. 2025, arXiv:2405.04383, MNRAS, 536, 2914, COMAP Galactic science I: observations of spinning dust emission at 30 GHz in dark clouds surrounding the λ-Orionis H II region

9. Dunne, D. A., et al. 2025, arXiv:2503.21743, Submitted to A&A, Three-Dimensional Stacking as a Line Intensity Mapping Statistic

BeyondPlanck and Cosmoglobe papers

10.

11. Watts, D. J., et al. 2024a, arXiv:2406.01491, arXiv e-prints, arXiv:2406.01491, Cosmoglobe DR2. II. CIB monopole measurements from COBE-DIRBE through global Bayesian analysis

12. Watts, D. J., et al. 2023a, arXiv:2202.11979, A&A, 675, A16, From BEYONDPLANCK to COSMOGLOBE: Preliminary WMAP Q-band analysis

13. Watts, D. J., et al. 2023b, arXiv:2303.08095, A&A, 679, A143, COSMOGLOBE DR1 results. I. Improved Wilkinson Microwave Anisotropy Probe maps through Bayesian end-to-end analysis

14. Andersen, K. J., et al. 2023, arXiv:2201.08188, A&A, 675, A13, BEYONDPLANCK. XIII. Intensity foreground sampling, degeneracies, and priors

15. Fuskeland, U., et al. 2023, arXiv:2302.05228, A&A, 676, A42, Tensor-to-scalar ratio forecasts for extended LiteBIRD frequency configurations

16. Eskilt, J. R., et al. 2023a, arXiv:2306.15511, A&A, 678, A169, COSMOGLOBE: Towards end-to-end CMB cosmological parameter estimation without likelihood approximations

17. Eskilt, J. R., et al. 2023b, arXiv:2305.02268, A&A, 679, A144, COSMOGLOBE DR1 results. II. Constraints on isotropic cosmic birefringence from reprocessed WMAP and Planck LFI data

18. BeyondPlanck Collaboration, et al. 2023, arXiv:2011.05609, A&A, 675, A1, BEYONDPLANCK. I. Global Bayesian analysis of the Planck Low Frequency Instrument data

19. Gerakakis, S., et al. 2023, arXiv:2205.11262, The Open Journal of Astrophysics, 6, 10, From BeyondPlanck to Cosmoglobe: Open Science, Reproducibility, and Data Longevity

Contents

Chapter 1

Cosmology

Cosmology is the science of the Universe itself, its history and future, including all its contents. Cosmology is concerned with the Universe at its largest scales, as opposed to astrophysics, which tries to answer the questions of how the Universe works locally on smaller and intermediate scales. However, while maturing into a modern high-precision science, it became clear that to understand the Universe as a whole, we also need to understand how it works locally on small scales as well. That is, there might be yet poorly understood physics of the small-scale Universe that couple to the very largest of scales.

Throughout this chapter, we will go through some of the history and concepts of modern cosmology. Unless otherwise stated, the general information on the cosmological formalism in the following is based on Dodelson (2003), Carroll (2019) and Particle Data Group Collaboration (2022) (Ch. 22.-29.).

1.1 Modern cosmology

When describing the evolution and structure of the Universe, modern cosmology relies on a few foundational principles. Under the assumption that the Universe on the largest scales is *isotropic* (i.e., looks equivalent in any direction) and using the *Copernican principle*, that the Universe has no special reference frame, we obtain the *Cosmological principle*: the Universe on large scales is both isotropic and homogeneous.

Combining this principle with the *theory of general relativity* (GR), derived by Einstein (1915), we find that space-time and the matter-energy content of the Universe interact with each other via the compact form

$$\mathsf{G}_{\mu\nu} + \Lambda \mathsf{g}_{\mu\nu} = 8\pi G \mathsf{T}_{\mu\nu}, \tag{1.1}$$

where G and Λ are the gravitational and cosmological constants, respectively. $\mathsf{G}_{\mu\nu}$ is the *Einstein tensor*, $\mathsf{g}_{\mu\nu}$ is the *metric tensor*, and $\mathsf{T}_{\mu\nu}$ is the *energy-momentum tensor*. The left- and right-hand sides of Eq. (1.1), respectively, describe the geometry and energy-matter content of space-time.

Shortly after the derivation of GR, Edwin Hubble and Georges Lemaître observed the recessional velocity of distant galaxies as a function of separation distance and deduced that the Universe seemed to expand with time (Lemaître, 1927; Hubble, 1929). This expansion of the Universe can be parameterized by a scale factor, $a(t)$, as a function of time, t, by

$$\frac{a(t)}{a_0} = \frac{\lambda_{\mathrm{obs}}}{\lambda_{\mathrm{em}}} = \frac{1}{1+z}, \tag{1.2}$$

where $a_0 = 1$ is the scale factor today, z is the redshift of the light, and λ_{obs} and λ_{em} the received and emitted wavelengths, respectively. Thus, when measuring light that has traveled through space over cosmological distances, we can deduce when it was emitted in the Universe's history by measuring its redshift.

Assuming the Universe follows the Cosmological principle, Einstein's field equations can then be solved to obtain the time evolution of the scale factor $a(t)$ through its expansion rate $H(t) = \frac{\dot{a}(t)}{a(t)}$;

$$H^2 = H_0^2 \left(\frac{\Omega_{\gamma,0}}{a^4} + \frac{\Omega_{c,0} + \Omega_{b,0}}{a^3} + \Omega_{\Lambda,0} \right). \tag{1.3}$$

This so-called *Friedmann equation* is parameterized by the expansion rate today, the *Hubble parameter*, $H_0 \approx 70\,\mathrm{km/s/Mpc}$, as well as the energy-density parameters $\Omega_{\gamma,0}$, $\Omega_{c,0}$, $\Omega_{b,0}$ and $\Omega_{\Lambda,0}$ of, respectively, *photons* (i.e., light), *cold dark matter* (CDM), *baryonic matter* (i.e., "normal" matter) and *dark energy* (Λ) at present time.

With this relatively simple equation, the evolution of the Universe on the very largest scales can be predicted. Furthermore, if the Universe is expanding, it could have started out with (near) infinite density from which all we know has emerged. Throughout the 20$^{\text{th}}$ and 21$^{\text{st}}$ centuries, this so-called *Big Bang theory* has been tested again and again by experimental evidence and has resulted in the current cosmological standard model, the ΛCDM model (Planck Collaboration I., 2020).

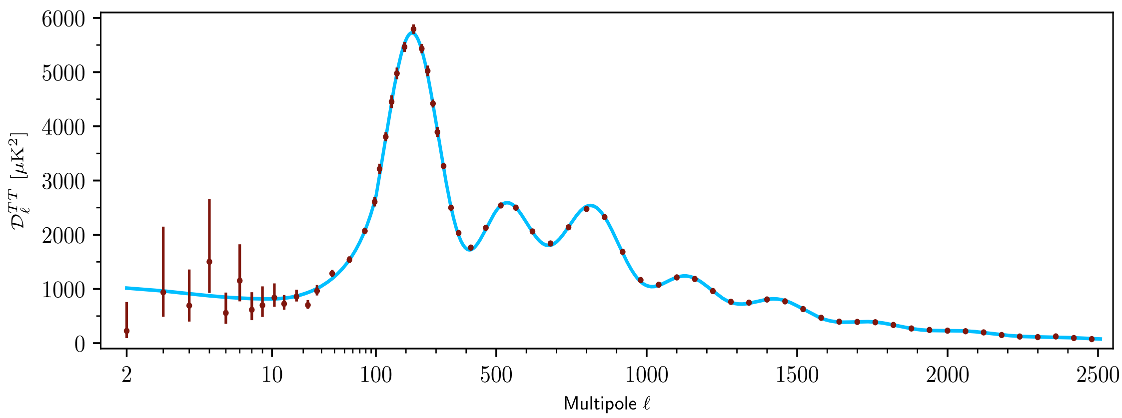


Figure 1: Planck CMB temperature power spectrum (red) and the best-fit ΛCDM prediction (blue). The power spectrum can be thought of as the variance of structures in the CMB temperature map seen in Fig. 4 (y-axis) at different scales (given by the multiples ℓ on the x-axis). See Sec. 3.6.1 for more on power spectra. Adapted from Planck Collaboration I. (2020).

The ΛCDM model has been immensely successful at describing the evolution of the Universe across its ~ 13.8 billion years of existence, and describes cosmological observations from Mpc to (particle) horizon scales (i.e., the distance a photon could have traveled since the beginning of time) with only six free parameters (Planck Collaboration I., 2020). An example of this striking agreement between the model and observables is shown in Fig. 1, showing the Planck temperature power spectrum (Planck Collaboration IV., 2020) and the best fit ΛCDM prediction. It is characterized by a matter-energy budget at the present day dominated by cold dark matter and a cosmological constant (dark energy). Together, these two components comprise around 95 % of the energy in the Universe today (Planck Collaboration I., 2020). The geometry of the ΛCDM model is spatially flat (i.e., parallel light rays forever remain parallel), while it is spatially expanding over time.

With its six parameters constrained to percent level accuracy and the immense explaining power and simplicity of the ΛCDM model, it has firmly cemented the Big Bang paradigm as the leading theory of the Universe (Planck Collaboration I., 2020). Despite its success, several big questions about the nature of our Universe remain to be answered. As we will later discuss, this thesis is trying to aid in filling this gap of knowledge by considering line intensity mapping to probe the poorly understood epochs of the Universe. However, before we do this, we briefly introduce the history of the Universe as we believe it to have happened.

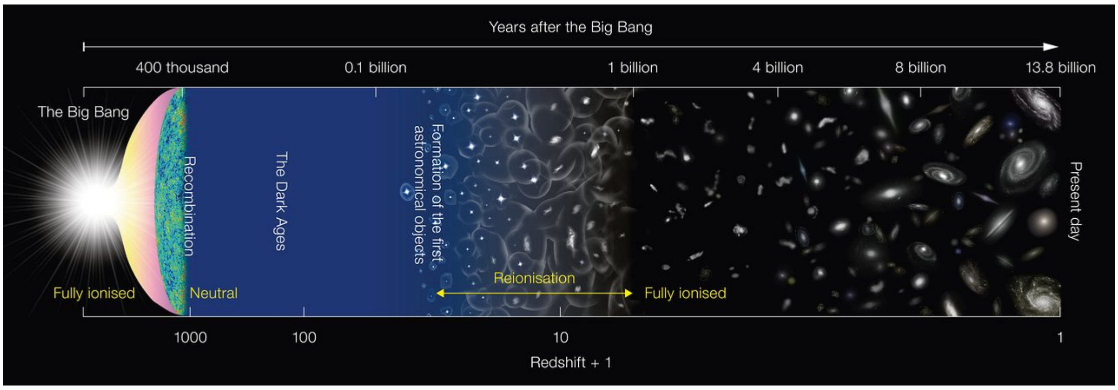


Figure 2: Schematic of the history of the Universe from the Big Bang, through the Epochs of Recombination, the Dark Ages, Reionization, Galaxy Assembly, until the present. Image credit: NAOJ

1.2 The history of the Universe in three acts

We can divide the evolution of the Universe into three main acts: the Universe at primordial times, the Universe at recombination, and the Universe from the Dark Ages to today. Figure 2 shows a schematic of the timeline of the Universe through its 13.8 billion-year existence. As context for our later discussion on line intensity mapping in Ch. 2, we will in the following briefly summarize the Universe's history in its three main parts. For a review of the history of the Universe, see Dodelson (2003), Schneider (2015) and Particle Data Group Collaboration (2022; Ch. 22.–29.).

1.2.1 The Universe at primordial times

As discussed earlier, we know that the Universe is expanding. This suggests that when going back to primordial times, the Universe must have been near infinitely hot and dense as the scale factor $a \to 0$. We do not know much about out Universe's origin as our current physical theories break down at those energy scales, nevertheless, we call this initial moment the Big Bang (Carroll, 2019; Particle Data Group Collaboration, 2022, Ch. 22).

We know from observations of the cosmic microwave background (CMB) that the Universe at early times was very hot, dense, and smooth, with density and temperature fluctuations from the mean of only $\delta \sim 10^{-5}$ (Planck Collaboration I., 2020). However, this poses an interesting question: Why are there small fluctuations rather than a completely smooth universe, and what seeded these anisotropies?

It turns out that these problems can all be solved by a brief yet extremely rapid period of expansion at early times (see Vazquez et al., 2018; Achúcarro et al., 2022; Particle Data Group Collaboration, 2022, Ch. 23., for reviews). This phase of exponential expansion right after the Big Bang, first proposed by Guth (1981), is called *inflation* and is currently one of the most active areas of CMB research. The cause of inflation is still a highly debated topic in cosmology. One hypothesis is that inflation was caused by a scalar field, the *inflaton* field, which caused the exponential expansion of the Universe and eventually decayed into the standard model particles of quantum physics, marking the end of cosmic inflation (Linde, 1982; Albrecht & Steinhardt, 1982; Kofman et al., 1994). As a consequence, it could blow up quantum fluctuations to cosmological scales and thus could be the very mechanism that gave rise to the anisotropies observed in the CMB (Mukhanov & Chibisov, 1981, 1982; Hawking, 1982; Guth & Pi, 1982; Starobinsky, 1982; Bardeen et al., 1983; Mukhanov, 1985).

Such an expansion phase would result in points at very large distances being causally connected, thereby solving the so-called *horizon problem*. The horizon problem arose when it was observed that scales larger than $\sim 0.5°$ had the same temperature, despite the large distance implying they should not have had time to interact and come to thermal equilibrium within the age of the Universe. Additionally, the Universe could start with any spatial geometry as a short inflation period would straighten out any curvature (just like the wrinkles in a deflated balloon disappear when inflating it). This thus solves the so-called *flatness problem* of cosmology where very special initial conditions would otherwise be required to explain the observed spatial flatness of the Universe. Cosmic inflation as a solution to the horizon and flatness problems was initially proposed by Guth (1981) Linde (1982), Albrecht & Steinhardt (1982), Kazanas (1980) and Sato (1981). Current estimates suggest that an expansion of 50–60 e-foldings (i.e., an expansion of e^x) over a fraction of a second would be enough to be consistent with observations (Komatsu et al., 2009; Planck Collaboration X., 2020; Particle Data Group Collaboration, 2022, Ch. 23.).

At the end of inflation, after the first particles and anti-particles had formed, the Universe was still hot and dense, and particles, anti-particles, and photons would be in an equilibrium of annihilation and creation. Not until the temperature

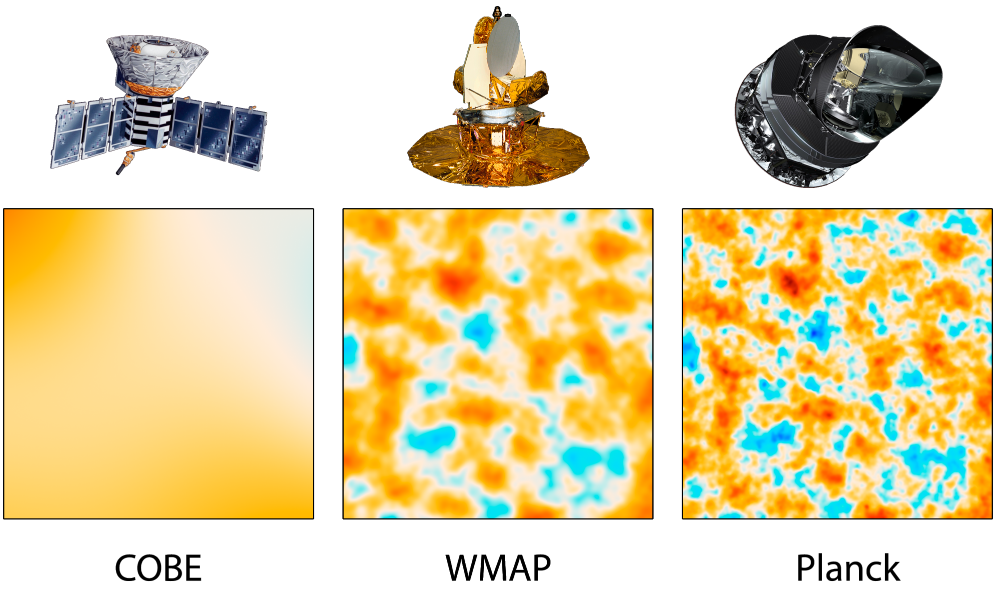


Figure 3: Three main CMB satellite missions to date with $10^\circ \times 10^\circ$ sections of their respective full-sky CMB maps. Image credit: NASA/JPL-Caltech/ESA.

of the Universe was below about $k_B T \sim 1\,\mathrm{MeV}$ was it cold enough for neutrinos to decouple and free stream. Atomic nuclei could not form until the temperature of the expanding Universe had reached $k_B T \sim 0.1\,\mathrm{MeV}$. This epoch is known as Big Bang nucleosynthesis (BBN), and the predicted relative abundances of light elements such as hydrogen, deuterium, and lithium, and predicted mass fractions of baryons, are additional great successes of the Big Bang theory (Planck Collaboration I., 2020; Particle Data Group Collaboration, 2022, Ch. 24.).

1.2.2 Recombination and the cosmic microwave background

In the second act of the Universe's evolution, we focus on its transition from a hot plasma to a neutral state. A telltale sign of a Big Bang universe that initially was much denser and warmer is an isotropic background radiation that follows a thermal blackbody spectrum. In such a model, the Universe would, shortly after the Big Bang, consist of a hot photon-matter plasma of photons scattering off free electrons. However, due to the Universe's expansion, the primordial plasma's temperature decreased until it was cold enough for the free electrons to bond with atomic nuclei, forming neutral atoms. This point, roughly 380.000 years after the Big Bang, is called *recombination*. The photons initially trapped by scattering on the free electrons could now freely travel through the Universe. Thus, any observer in the Universe will see a shell of cosmic photons of radius corresponding to the distance the light has traveled since recombination. This shell is often referred to as the *last scattering surface* (LSS). This radiation from the afterglow of the Big Bang was first discovered accidentally by Penzias & Wilson (1965) in the microwave spectrum as an anomalous $\sim 3\,\mathrm{K}$ excess in antenna temperature. We refer the interested reader to Ch. 22. and 29. of Particle Data Group Collaboration

(2022) for an extensive review on recombination. Today, about half a century later, we have several high-precision satellite measurements of the CMB. Starting with the Cosmic Background Explorer (COBE; Mather et al., 1990, 1994), launched in 1989 into a low-Earth orbit, it was shown with exquisite sensitivity that the CMB is an almost perfect blackbody (Fixsen et al., 1996) and that the CMB is very smooth with fluctuations of only $\delta \sim 10^{-5}$ from the mean (Smoot et al., 1992; Bennett et al., 1994, 1996).

Although COBE measured the CMB anisotropies, its angular resolution was relatively poor. This was improved by the Wilkinson Microwave Anisotropy Probe (WMAP) in 2001 (Bennett et al., 2003) and the Planck satellite in 2009 (Planck Collaboration I., 2011). They were both launched to the second Sun-Earth Lagrange point and produced two important legacy CMB anisotropy maps. The progress in CMB maps from COBE, WMAP to Planck, and the full-sky CMB map of the Planck DR3 release (Planck Collaboration VI., 2020), can be seen in Figs. 3 and 4. Although the CMB anisotropies are mere $\sim 10^{-5}$ fluctuations from the mean temperature, they are thought to be the very seeds from which more advanced structures like galaxies and galaxy clusters formed.

As CMB experiments like Planck and WMAP have measured the temperature fluctuations in great detail, the focus of current and future CMB surveys has transitioned to more subtle effects such as CMB polarization. The CMB polarization can be broken down into two components, the E- and B-mode polarization, see Fig. 5. The E-mode pattern has already been measured by Planck (Planck Collaboration I., 2020). It is caused by the scattering of photons in local temperature quadrupoles due to pressure waves in the photon-baryon plasma at recombination. One of the predictions of the epoch of cosmic inflation we discussed in Sec. 1.2.1 is the existence of primordial gravitational waves. These are currently one of the only primordial physical mechanisms that would form a B*-mode* polarization pattern in the CMB(Hu & White, 1997; Zaldarriaga & Seljak, 1997; Dodelson, 2003; Kamionkowski & Kovetz, 2016). As such, B-modes could be a direct window back to inflation and help probe the Universe at its very earliest moments (Seljak & Zaldarriaga, 1997; Kamionkowski et al., 1997; Kamionkowski & Kovetz, 2016). Additionally, inflation predicts the existence of a non-Gaussian distribution of anisotropies in the CMB and subsequent structures that formed from it (see Planck Collaboration IX., 2020, for Planck constraints on CMB non-Gaussianities).

1.2.3 The maturing Universe – from recombination to structure formation

We have reached the third and longest (in terms of cosmic time) act of the history of the Universe. It starts with recombination and progresses to the current age. In the following, we summarize the highlights of this phase of the Universe's history (for more details, see, for instance, textbooks Dodelson (2003), Schneider (2015) and (Carroll, 2019), as well as reviews Ch. 22., 25. and 29. of (Particle Data Group Collaboration, 2022)).

At recombination, when the temperature of the photon-baryon plasma dropped

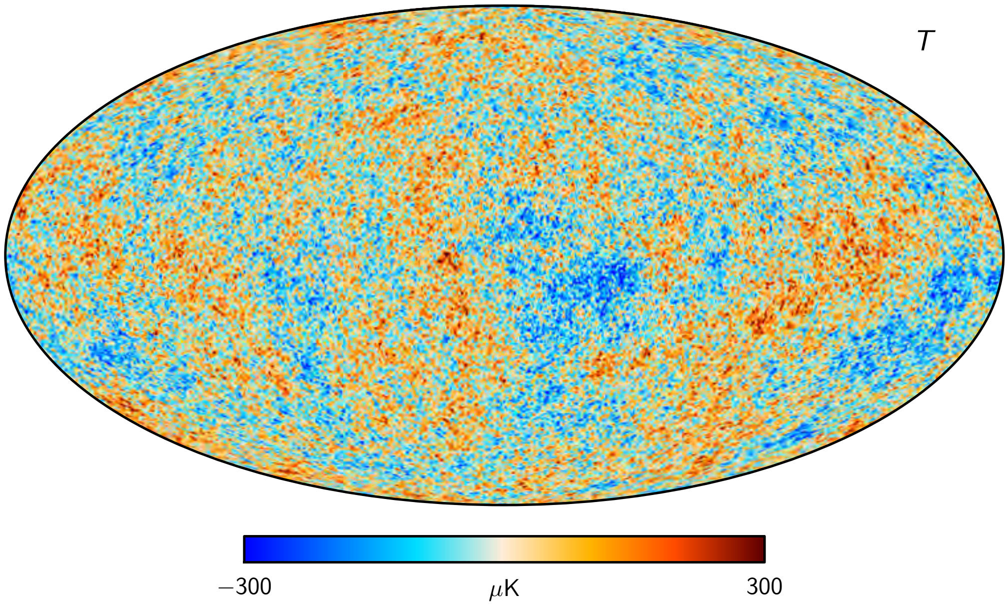


Figure 4: Full-sky CMB temperature anisotropy full-sky map from the Planck DR3 Commander pipeline (Planck Collaboration IV., 2020). The figure is adapted from Planck Collaboration IV. (2020).

below the ionization energy of hydrogen, about 13.6 eV, breaking the coupling between baryons and photons as neutral atoms formed. Once the neutral atoms form, the photons can free-stream through the Universe, and the baryons are free to clump together. The dark matter is theorized to have decoupled from the photon bath much earlier than the regular matter and, as such, is able to clump much earlier. This provides the necessary gravitational wells to allow the baryonic matter to clump and form the structure we see today. This provides some of the best evidence for dark matter that we have today (Dodelson, 2003; Carroll, 2019).

For several hundred million years, during the so-called *Dark Ages*, the Universe was filled with cold neutral gas, with the only emitted radiation being from the 21 cm hydrogen line. The cold neutral gas eventually clumped together enough to form the first stars and galaxies (see, e.g., Hashimoto et al., 2018; Willis et al., 2020). This period called the *Cosmic Dawn*, marked the beginning of the *Epoch of Reionization* (EoR) in which the radiation of stars and galaxies heated the previously cold neutral Universe to once again become ionized. Observations of spectra of distant quasars[1] (e.g., Gunn & Peterson, 1965; Becker et al., 2001) and estimates of the optical depth to reionization by Planck (Planck Collaboration I., 2020) suggest this transition happened around redshift $z \sim 6 - 10$ (see Eq. (1.2)), and at $z \sim 6$ the Universe was entirely ionized. Though current evidence suggests that reionization occurred fast and late and was initially driven by high-mass stars in small galaxies, there are significant uncertainties regarding the details of reionization, and it remains an active field of research (Planck Collaboration I., 2020; Kovetz et al., 2017).

[1]Quasar stands for "quasi-stellar object" (QSO). They are active galactic nuclei that glow bright due to the accretion of matter onto a supermassive black hole. See Schneider (2015).

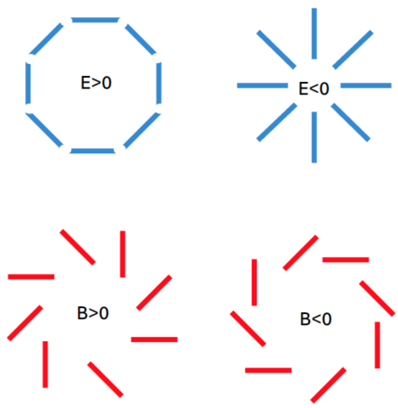


Figure 5: Examples of E- and B-mode polarization patterns (*top* and *bottom*, respectively). The figure is taken from Rahimi & Reichardt (2024).

In surveys of the galaxy distribution on large scales, it can be seen that matter clumped to form structure on some characteristic scales corresponding to roughly 150 cMpc (Eisenstein et al., 2005; Cole et al., 2005). This preferential clustering is caused by the so-called baryonic acoustic oscillations (BAO). The scale at which BAOs occur corresponds to the distance a photon-baryon pressure wave could have traveled up until they decoupled. The baryonic structure was frozen in place while the photons traveled onward. Hence, BAOs form an important standard ruler that can be calibrated from the first acoustic peak of the CMB power spectrum seen in Fig. 1 to measure cosmological distances and the cosmic expansion history (Dodelson, 2003; Particle Data Group Collaboration, 2022, Ch. 22.).

From redshift $z \sim 8$ to 3, the star formation rate steadily grew until it peaked somewhere around $z \sim 2$, after which it again declined until the present day. During the epoch of peak cosmic star formation, around $z \sim 2$, also referred to by the *Epoch of Galaxy Assembly* (EoGA) or *Cosmic Noon*, the star formation rate density (SFRD) was around ten times larger than at the present (see Madau & Dickinson, 2014, and references therein, for extensive review of the cosmic star formation history).

The galaxies that formed during these times up until today collectively emit a glowing cosmic background (CB) of electromagnetic radiation, not too dissimilar from the CMB. This radiation, the *extragalactic background light* (EBL), can be decomposed into different components such as the cosmic infrared, optical, and line emission background (CIB, COB, and CLB, respectively), the latter of which we will discuss more in Ch. 2 about mapping the CLB with line intensity mapping. As we will discuss later in Ch. 3, mapping the cosmic carbon monoxide (CO) line emission fluctuations is the primary goal of the COMAP experiment, the main subject of this thesis. The CMB was emitted during a relatively short amount of cosmic time and hence appears to us as a 2D projected last scattering surface. In contrast, the EBL is the product of all cumulative emissions from the 3D structure of stars and galaxies in the Universe since they started to form at Cosmic Dawn until today (see Lagache et al., 2005; Hill et al., 2018; Mashian et al., 2016, for more information about EBL).

Going further through cosmic time, we eventually reach the present. This epoch is characterized by an exponential expansion of the Universe, as directly observed by measuring the distance to distant galaxies using supernovae type Ia used as *standard(izable) candles* (Riess et al., 1998; Perlmutter et al., 1999; Riess et al., 2022a,b), but which can also be inferred from Planck observations of the CMB power spectrum (Planck Collaboration I., 2020). The exponential expansion is thought to be caused by some negative pressure component with constant energy density, often called *dark energy* or *cosmological constant*. Though we do not know precisely what the *cosmological constant* is, it is a phenomenological component needed to explain this expansion in the standard ΛCDM model mentioned in Sec. 1.1.

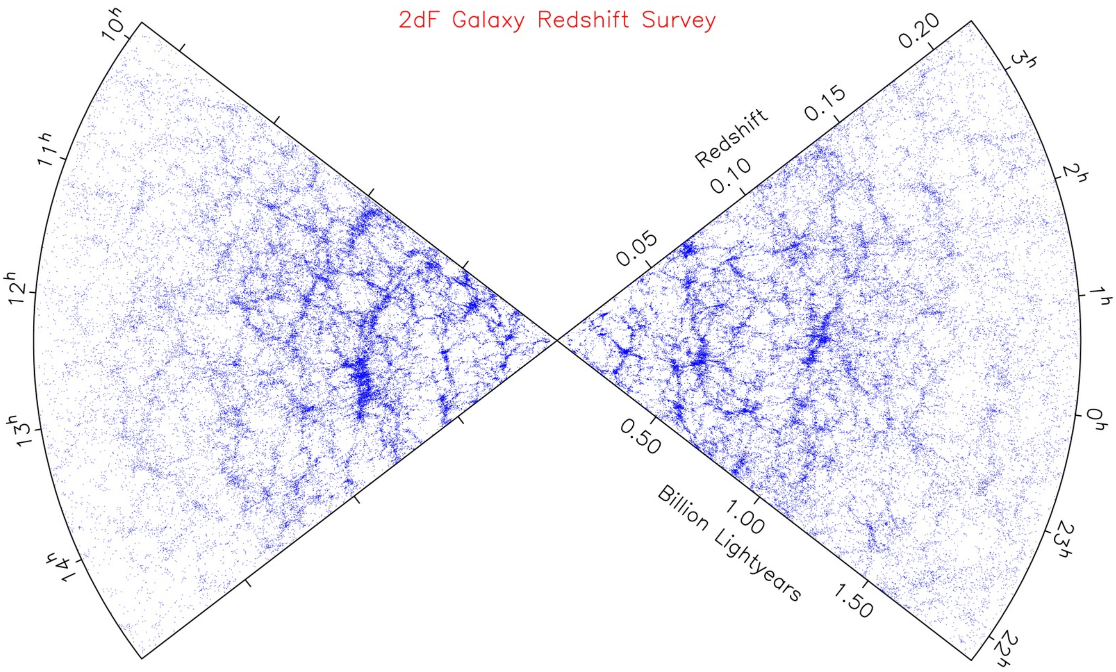


Figure 6: Slice through the large-scale structure as mapped by the 2dF Galaxy Redshift Survey (2dFGRS). Each blue dot represents a galaxy. Image credit: 2dFGRS.

The physics of structure formation at late times is highly complex as matter perturbations have collapsed into highly nonlinear structures governed by complicated interactions. Most of our direct knowledge about galaxies and the large-scale structure at low redshift has been obtained through galaxy surveys like the Sloan Digital Sky Survey (SDSS; Eisenstein et al., 2005; Alam et al., 2021; Abdurro'uf et al., 2022), 2dF Galaxy Redshift Survey (2dFGRS; Colless et al., 2001; Cole et al., 2005), the 6dF Galaxy Survey (6dFGS; Jones et al., 2009; Beutler et al., 2011), and the WiggleZ Dark Energy Survey (Drinkwater et al., 2010; Blake et al., 2012). These have mapped the large-scale structure by measuring the position and brightness of large numbers of individual galaxies and quasars, as shown in Fig. 6, which shows an example of a 2dFGRS large-scale structure survey. However, as we will discuss in the next chapter, galaxy surveys have their limitations, and many unanswered questions about the nature of astrophysics and cosmology remain to be answered by future surveys (such as COMAP and other line intensity mappers).

Chapter 2

Line Intensity Mapping: generalizing a CMB survey to map 3D volumes

As discussed in the previous chapter, we know a lot about the evolution and contents of the Universe from, e.g., CMB and galaxy surveys. However, there are still many questions to which we do not know the answers. In particular, the periods between recombination and the present, i.e., the Dark Ages, Cosmic Dawn, EoR and the EoGA (see Fig. 2), are data-starved research fields. To overcome this lack of knowledge, a promising and relatively new field called *line intensity mapping* (LIM) has been developed over the last decades and is starting to gather data. As a contribution to this field, the work presented in the Papers I-V are important milestones. COMAP is one of the leading LIM experiments and will be discussed further in Ch. 3. In this chapter, we will explore what LIM is, some of its scientific goals, and experiments that contribute to the field.

Most of the information presented in this chapter is based on the reviews of Kovetz et al. (2017) and Bernal & Kovetz (2022), and we refer the interested reader to their work, as well as references therein, for more details on the field in general.

2.1 Line intensity mapping

Line intensity mapping is a field within observational astrophysics and cosmology designed to fill the gap of detailed direct observations of the EoR (Madau et al., 1997; Battye et al., 2004; Peterson et al., 2006; Loeb & Wyithe, 2008). It aims to map the Universe volumetrically from the current era back to recombination when the Universe becomes opaque.

To do this, a LIM survey will collect all photons of some atomic or molecular emission line from a given volume of the Universe, whether emitted from a bright point source like a galaxy cluster or diffuse gas and dwarf galaxies. Furthermore, by measuring the frequency of the received light, we can infer the cosmic redshift and, hence, the line-of-sight look-back distance the photons have traveled.

An example of a simulated LIM survey in a CO rotational and the [C II] ionized carbon emission lines can be seen in Fig. 7. As mentioned in Ch. 1, the 3D large-scale structure has historically been systematically probed by galaxy redshift surveys. They do this by probing only the brightest point sources of the

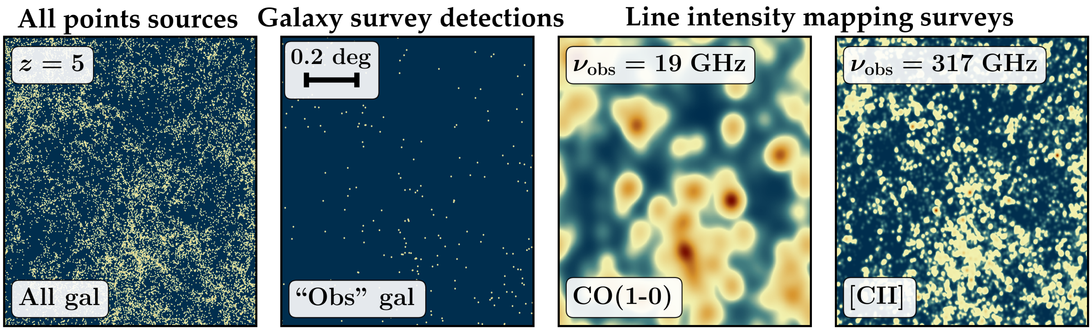


Figure 7: Simulated comparison between mapping large-scale structure with galaxy and line intensity mapping surveys. From left to right, the panels show a simulated galaxy distribution at redshift $z = 5$, the galaxies as observed by galaxy surveys, and respectively an unbiased CO and [C II] line intensity mapping survey. Adapted from Bernal & Kovetz (2022).

structure, represented by the yellow dots seen in the two left panels of Fig. 7, and hence provide a biased view of the large-scale structure. That is, there might be a significant amount of faint and diffuse matter below the detection threshold of a galaxy survey. Meanwhile, LIM surveys trade resolution power for a direct measurement of the large-scale structure, as seen in the two right panels of Fig. 7. This is achieved by collecting more photons in the same amount of integration time but from a larger area. As such, galaxy surveys are better suited to explore the physics within galaxies, while LIM's unbiased maps of cosmic structures are superior at measuring the statistical properties of the probed field. Line intensity surveys are, therefore, both quicker and cheaper at mapping out the matter in the Universe as compared to galaxy surveys.

Intensity mapping maps the cosmic line emission background (CLB) fluctuations in 3D space by using the spectral information of different emission lines. This is unlike a CMB or EBL survey, which can only observe the 2D projected intensity. In other words, one can think of LIM as a CMB survey with a high-resolution spectrometer attached to it. Many of the current line intensity mapping experiments build on the expertise of the CMB community because the observational techniques and data analysis of LIM are very similar to that of a classic cosmic background experiment.

Although LIM will greatly complement existing cosmology and astrophysics experiments, it is important to note that it also has its challenges. Firstly, as mentioned earlier, LIM cannot resolve individual galaxies and, therefore, cannot give detailed insight into individual galaxies, their morphology, etc. Secondly, emission lines from a different spectral line can be redshifted into the experimental bandpass. This is known as the interloper problem, illustrated in Fig. 8. Some proposed solutions to this problem (as summarized by Kovetz et al., 2017) are cross-correlations between different LIM surveys targeting the same redshift with different emission lines and bandpasses (Visbal & Loeb, 2010; Gong et al., 2012, 2014; Roy et al., 2024; Fronenberg & Liu, 2024), targeted masking of bright known foreground sources (Visbal et al., 2011; Gong et al., 2014; Breysse et al., 2015; Silva et al., 2013, 2015, 2021; Sun et al., 2018), as well as using the Alcock-Paczyński

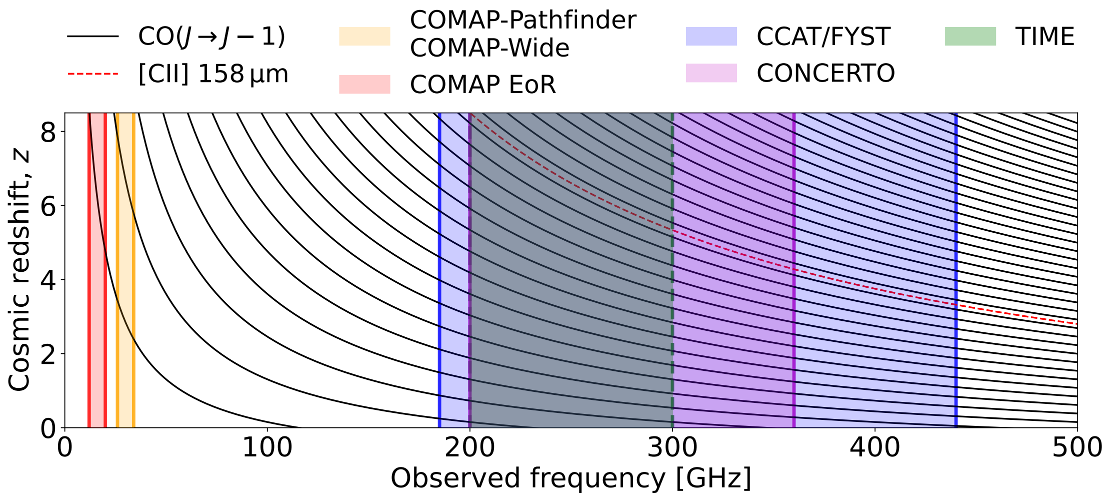


Figure 8: Overview of several $CO(J \rightarrow J{-}1)$ rotational lines, and the 158 μm [C II] line as a function of redshift and observed frequency. Experiment pass-bands are shown as shaded regions.

effect to separate dominant foreground interlopers at the wrong redshift (Lidz & Taylor, 2016; Cheng et al., 2016). Lastly, LIM requires a high dynamic range as the target signal is extremely weak, while the foregrounds and systematics are orders of magnitude brighter. Thus, LIM observations are relatively easy to perform due to the fast mapping speed and lack of angular resolution but challenging in post-processing due to the dynamic range problem (see more discussion on LIM's advantages and challenges in the review by Kovetz et al., 2017).

2.2 Science goals and experiments

As mentioned, LIM, as originally proposed, was mainly focused on mapping the Universe in the 21 cm neutral hydrogen line to infer the physics of the EoR and Cosmic Dawn. However, since those early days of LIM numerous different emission lines, such as CO rotation lines, the [C II] fine structure line, Lyα, Hα, Hβ, [O II] and [O III], have been proposed to probe a wide range of scales, physical environments and epochs (Kewley et al., 2019). Today, numerous LIM experiments are ongoing, funded, or are being proposed. In Fig. 10, we show an overview of the survey area of the most important proposed or ongoing experiments grouped by their line of interest.

We can divide the LIM science goals and the corresponding experiments into roughly three groups (with some overlap): the EoR, the EoGA, and large-scale structure and cosmology. In the following, we will summarize the main aspects of the three main LIM science regimes and mention some experiments within each group.

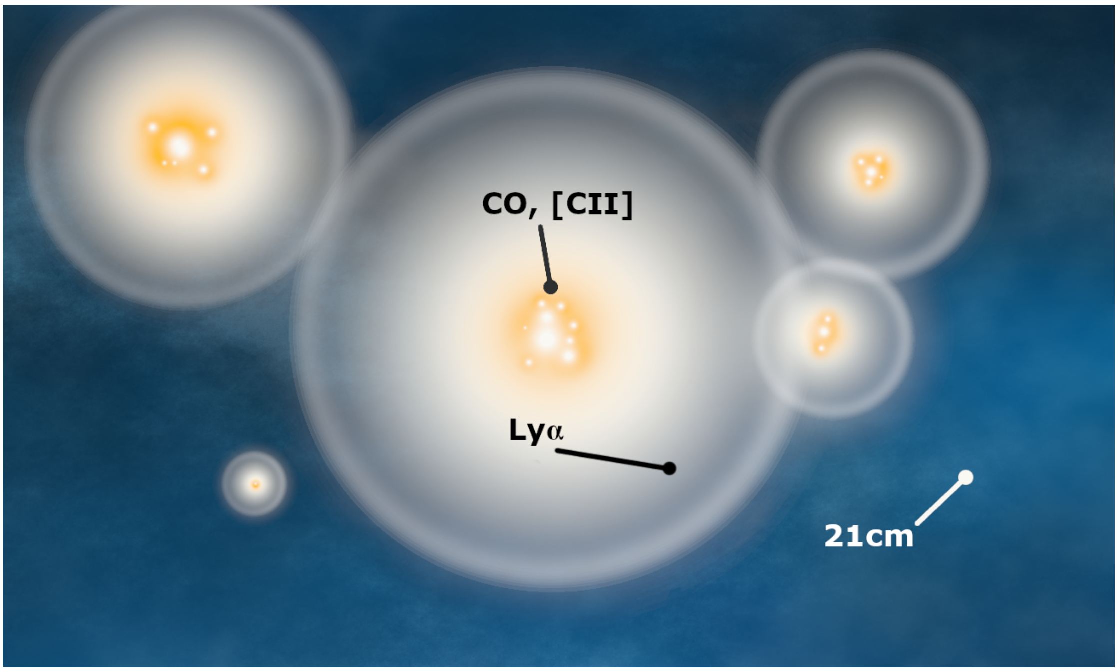


Figure 9: Interplay between different emission lines in the IGM and ISM around the EoR. The neutral IGM (in blue) is traced by the hydrogen 21 cm line, while CO traces star-forming galaxy clusters (bright yellow) and [CII]. The clusters create ionization bubbles (diffuse yellow-white) traced by the Lyα line.

2.2.1 Epoch of Reionization and the Dark Ages

As mentioned earlier in Ch. 1, we know that the Universe transitioned from neutral to ionized around redshift $z \approx 6$–15. For instance, evidence for this period is the smoothing of the CMB anisotropies due to CMB photons scattering off free electrons at reionization (Planck Collaboration VI., 2020). We also observe a Gunn-Peterson trough in the spectra of distant quasars (see, e.g., Becker et al., 2001; Bouwens et al., 2015) suggesting the inter-galactic medium went from a neutral to ionized hydrogen state. Currently, it is thought that the Universe was reionized relatively late and fully ionized by $z \sim 6$, by massive low-metallicity stars in small galaxies (Robertson et al., 2013, 2015). However, details about the exact processes that ionized the Universe are largely unknown. Was reionization primarily driven by massive population III stars[1], supernova feedback, or quasars in large or small galaxies? How long did reionization last? What does the topology of the reionization bubbles in the Universe look like?

Line intensity mapping aims to answer these questions by making maps of the large-scale structure from Cosmic Dawn to the end of the EoR. In fact, most initial efforts of LIM pertained to the study of the EoR (Morales & Wyithe, 2010; Furlanetto et al., 2006). Because of the unbiased nature of LIM, the effect of all emitters on reionization is captured, both from bright large galaxies and quasars, as well as smaller dwarf galaxies. Using the neutral 21 cm (H I) line, the neutral

[1]Population III stars refer to the first generation of massive and ultra metal-poor stars. For comparison, the Sun is a Population I star, while the oldest detected metal-poor stars are called Population II stars. See, e.g., Bodenheimer (2011); Schneider (2015).

intergalactic and interstellar media (IGM and ISM, respectively) can be mapped through the EoR and potentially back to the CMB. The reionization bubbles around ionizing UV emitters will appear as dark imprints in the H I map (Kovetz et al., 2017). In addition, the CO rotational lines and the [C II] fine structure 158 μm line of ionized carbon map the molecular clouds and dusty star-forming medium within galaxy clusters. These CO and [C II] maps will trace the center of the ionization bubbles. The bubbles themselves are traced by ionizing UV radiation and can be mapped using the Lyα hydrogen line.

This interplay between the different emission lines, as illustrated in Fig. 9, is a powerful probe of the physics of reionization on a wide range of scales. Measuring the same cosmic volumes with two or more different emission lines can provide an effective way to cancel systematic effects and noise of different surveys and help to confirm or rule out any falsely claimed detections (Kovetz et al., 2017).

Examples of surveys that work on mapping the EoR and Cosmic Dawn in H I are the interferometric experiments HERA (Abdurashidova et al., 2022) and SKA-LOW (Santos et al., 2015) at, respectively, $z \sim$ 5–27 and $z \sim$ 3–7. The ground based single-dish experiments CCAT-Prime/FYST ($z \sim$ 3.3–9.3, CCAT-Prime Collaboration et al., 2023), TIME ($z \sim$ 5–9, Staniszewski et al., 2014) and CONCERTO ($z \sim$ 4.3–8.5, Fasano et al., 2024; Van Cuyck et al., 2023) target [C II] and the future COMAP-EoR survey (Cleary et al., 2022; Breysse et al., 2022) targets CO ($z \sim$ 5–8) to constrain how star formation drove reionization. Lastly, the recently launched satellite mission SPHEREx, targeting the Lyα line at redshifts of $z \sim$ 5.2–8, will provide measurements of the ionized IGM and star formation during reionization (Doré et al., 2018). As seen in Fig. 8, the [C II] surveys generally also have several CO rotational lines as interloper foregrounds and can thus also perform LIM on the EoGA. An overview of some of the survey areas on the sky is shown in Fig. 10.

2.2.2 Epoch of Galaxy Assembly

Around half of today's stellar mass was formed at high redshifts $z \gtrsim 1.3$ in the EoGA, and just about 1 % of the stars were formed during the EoR (Madau & Dickinson, 2014). Consequently, to understand star formation processes, it is important to observe the star-forming galaxies at the EoGA $z \sim$ 2–5.

Our knowledge about high-redshift star formation stems primarily from bright individual galaxies, like the ones detected by Huynh et al. (2017), usually observed in UV, optical, or IR by galaxy surveys (Kovetz et al., 2017). Meanwhile, much of the star formation takes place in faint and cold molecular clouds in small galaxies. It is traced by CO rotational lines and the [C II] 158 μm fine structure line, which become exceedingly difficult to observe individually at high redshift (Hollenbach & Tielens, 1999; Kovetz et al., 2017; Lagache et al., 2018). Line intensity mapping, which collects the aggregate emission from all faint sources, is ideally suited for this task and can probe the star formation history without high-luminosity bias. Using empirical scaling relations, one can then connect the, e.g., CO, emission to the cosmic star formation rate density (SFRD) across cosmic time (see, e.g., Li et al., 2016; Chung et al., 2022a). We will present some more examples of this later in Sec. 3.6.5.

Among the experiments targeting the EoGA we have the single-dish surveys COMAP-Pathfinder (the main subject of this thesis, discussed in greater detail later in Ch. 3) mapping CO(1-0) at $z \sim 2$–3 (Cleary et al., 2022) and the upcoming SPT-SLIM, which will map several CO rotational lines in the range $z \sim 0.5$–2 (Karkare et al., 2022). We also have the interferometric CO Power Spectrum Survey (COPSS) at $z - 2.3$–3.3 (Keating et al., 2015, 2016) and mmIME $z \sim 1$–5 (Keating et al., 2020). Other experiments are the balloon-borne experiments EXCLAIM (Ade et al., 2020; Switzer et al., 2021) and TIM (previously called STARFIRE Vieira et al., 2020) that both target [C II] at, respectively, $z \sim 0$–3.5 and $z \sim 0.52$–1.67. The previously mentioned satellite mission SPHEREx will target Hα, Hβ, [O II] and [O III] during the EoGA at $z \sim 0.1$–5 to constrain star formation (Doré et al., 2018). The HETDEX experiment, originally designed as a Lyα galaxy survey targeting $z \sim 1.88$–3.52, can also be used as a LIM survey (Kovetz et al., 2017; Gebhardt et al., 2021). In Fig. 10, we see an overview of the conducted or planned survey areas on the sky. The small patch sizes of many EoGA surveys indicate that the experiments are in their pathfinder phase to demonstrate the first detections of their respective targeted lines before the field sizes are increased to cover more large-scale modes in the sky.

2.2.3 Large-scale structure and cosmology

Line intensity mapping can also be used as a powerful probe of cosmology by mapping the large-scale structure at low redshifts. As the fluctuations in the CMB are the very seeds for structure formation in later epochs of the Universe, one can access much of the same cosmological information at a later time by mapping out large linear scales (Kovetz et al., 2017).

For instance, one of the problems of research fields, such as primordial non-Gaussianities, is cosmic variance. That is, for each scale, there is only a limited number of possible ways to arrange structures. This gives rise to an intrinsic sample variance, called *cosmic variance*, that becomes larger for large-scale modes. When looking for primordial non-Gaussianities in the CMB, we are limited by the intrinsic uncertainty of cosmic variance due to having access to only a limited number of modes on a 2D projected surface. Meanwhile, a LIM survey will produce tomographic maps of large-scale structures that were seeded by the same fluctuations observed in the CMB. In such a 3D map, the cosmic variance is lower because there is access to more modes (Knox, 1995; Oxholm & Switzer, 2021). As such, fields like inflationary cosmology can greatly benefit from large-scale LIM surveys.

Other areas of interest within cosmology in which LIM can be highly complementary are, for instance, constraining the nature of dark matter and dark energy, modifications of standard ΛCDM, and general relativity. For example, LIM can map the evolution of large-scale structures across the transition from matter to dark energy domination. This is thus a powerful probe of the nature of dark energy (Kovetz et al., 2017).

In addition, LIM can be used to obtain accurate BAO measurements across cosmic time. With continuous standard ruler measurements from BAOs across time, we can get accurate information about the cosmic expansion history and angular diameter distance estimates up to high redshift. As such, LIM might provide a pathway to alleviate the tension between high- and low-redshift estimates of H_0 from, respectively, the CMB and supernova type Ia surveys (see Kamionkowski & Riess, 2023, for review on Hubble tension).

Among LIM surveys probing cosmology by mapping large-scale structure, we find experiments such as CHIME (Amiri et al., 2024; CHIME Collaboration et al., 2022), HIRAX (Crichton et al., 2022), SKA-MID and its precursor MeerKAT (Paul

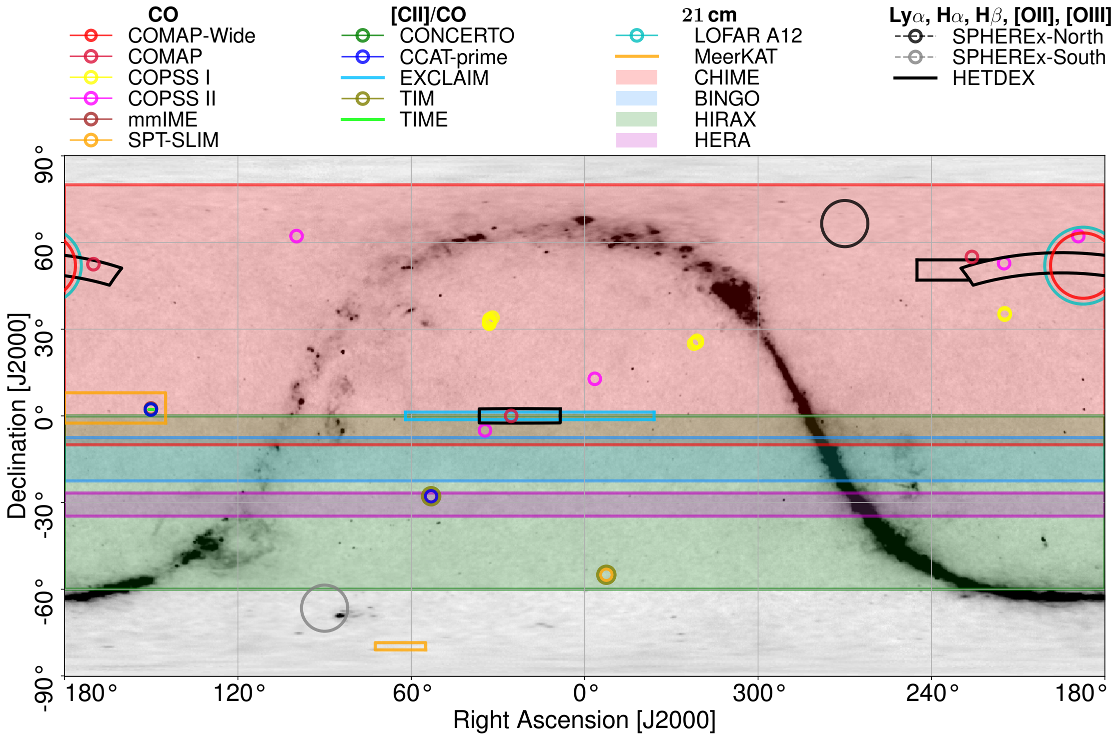


Figure 10: Overview of line intensity mapping fields across the sky from the literature: *COMAP* (Cleary et al., 2022), *COPSS I/II* (Keating et al., 2015, 2016), *mmIME* (Keating et al., 2020), *EXCLAIM* (Switzer et al., 2021), *CCAT-Prime/FYST* (CCAT-Prime Collaboration et al., 2023), *CONCERTO* (Fasano et al., 2024; Van Cuyck et al., 2023), *TIM* (Vieira et al., 2020; Agrawal et al., 2024), *HETDEX* (Gebhardt et al., 2021), *SPHEREx* (Doré et al., 2018), *CHIME* (Amiri et al., 2024; CHIME Collaboration et al., 2022), *BINGO* (Wuensche et al., 2021), *MeerKAT* (Paul et al., 2023; Cunnington et al., 2023; Mauch et al., 2020), *HERA* (Abdurashidova et al., 2022), *HIRAX* (Crichton et al., 2022), *TIME* (Staniszewski et al., 2014; Crites & Lau, 2024) and *SPT-SLIM* (Karkare et al., 2022; Zebrowski & Stover, 2024). The *COMAP-Wide* and *LOFAR* fields, which are completely overlapping due to the collaborative effort of the two surveys, are preliminary. Fields below 5° in radial size are plotted as 5° markers to make them visible on the plot. The experiments are grouped according to their emission lines of interest. The background shows the Planck LFI 30 GHz frequency map from the Planck Legacy Archive[2](Planck Collaboration I., 2020).

[2]Planck Legacy Archive: https://pla.esac.esa.int/pla/#maps

et al., 2023; Cunnington et al., 2023; Mauch et al., 2020), and BINGO (Wuensche et al., 2021). These target the H I line at redshifts of $z \sim 0.8$–2.5 (CHIME and HIRAX), $z \sim 0$–3 (SKA-MID) and $z \sim 0.13 - 0.48$ (BINGO) (Kovetz et al., 2017). In Fig. 10, we can see the survey areas of the different experiments. Note how interferometric 21 cm experiments targeting large-scale structures at low redshifts probe larger fields on the sky than typical single-dish surveys mapping the EoR and EoGA with other lines.

Chapter 3

The Carbon monOxide Mapping Array Project

Now that we have laid down the broader context of cosmology and how line intensity mapping can complement it, we can introduce the Carbon monOxide Mapping Array Project (COMAP). The COMAP experiment is the main focus of Papers I-V and, therefore, most of the work presented in this thesis. Specifically, the work presented in these papers can further be divided into the Early Science results (ES) (Papers I-II) and the Season 2 (S2) Results (Papers III-V). These contain the work of two successive data releases, one from 2022 presenting the state-of-the-art LIM pipeline and results and the other from 2024 with updated and significantly improved methodology and results.

We start with introducing the survey and instrument itself and subsequently go through the data model, systematics effects, and data analysis presented in Papers I-V roughly in the same order as the data are processed in the pipeline (i.e., low- and high-level analysis, and then modeling and inference).

3.1 The COMAP experiment

The COMAP experiment is a LIM experiment that aims to map the large-scale distribution of diffuse star-forming environment in the EoR and EoGA (Cleary et al., 2022). This is done by measuring the well-studied rotational lines of CO, the second most abundant molecule after H_2, that directly traces the molecular clouds in which stars form (Schulz, 2012). These emit (sub-)millimeter radiation at equally spaced frequencies of 115.27 GHz when jumping between rotational states and are therefore often called the *CO emission ladder* (Demtroder, 2010). The COMAP survey is built to utilize this fact and designed so that multiple CO rotational lines can be measured with the same experimental bandpass. Specifically, the survey will be conducted in several phases briefly summarized in the following, and in Fig. 11 showing the frequency-redshift coverage of COMAP and its phases together with two overlapping surveys COMAP will work within cross-correlation.

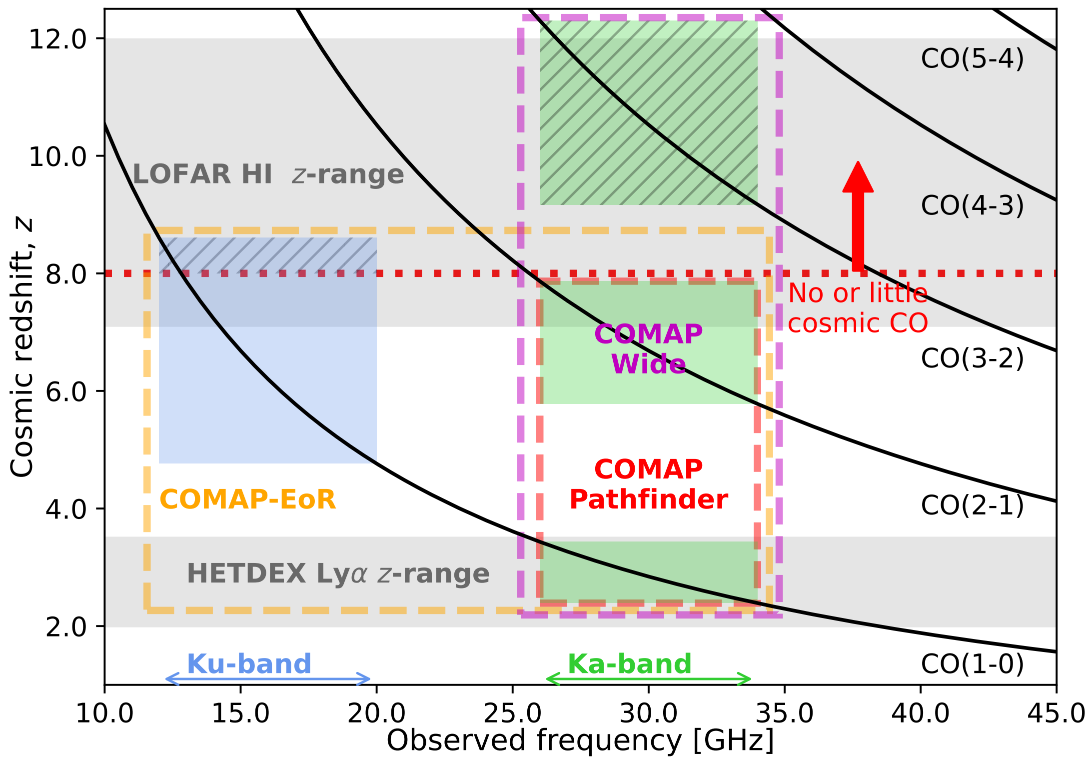


Figure 11: Frequency-redshift ranges covered by the COMAP-Pathfinder, COMAP-EoR, and COMAP-Wide phases of the experiment, across both the 30 GHz Ka- and 18 GHz Ku-bands. The redshift ranges of the overlapping LOFAR H I and HETDEX Lyα surveys are shown as gray bands. Redshifted CO($\mathrm{J} \rightarrow \mathrm{J}-1$) are indicated as black lines. Above the red dotted line at $z \sim 8$, there is expected to be a diminishing amount of cosmic CO (Barkana & Loeb, 2001), and therefore, we hatch corresponding COMAP CO-line overlap regions.

3.1.1 The COMAP-Pathfinder instrument

The ongoing five-year COMAP-Pathfinder, the main subject of this thesis, is the first phase of the COMAP survey. The Pathfinder aims to map the CO(1–0) 115.27 GHz and CO(2–1) 230.54 GHz (as well as possibly CO(3–2) 345.81 GHz) lines at frequencies between 26–34 GHz. These correspond, respectively, to redshifts $z = 2.4$–3.4 shortly before peak cosmic star formation at the EoGA, and $z = 6$–8 (and $z = 10$–12) in the EoR (see Fig. 11, though note that there is not expected to be much cosmic CO emission beyond $z \sim 8$; Barkana & Loeb, 2001). By mapping 12 deg^2 on the sky, the COMAP-Pathfinder is the first survey that is able to place direct constraints on the 3D clustering of CO(1–0) and is hence ideally suited to constrain global properties of the star-forming galaxies at the EoGA. The COMAP-Pathfinder is thus a test bed for technology and will demonstrate the possibility of CO line intensity mapping at the clustering scale (Cleary et al., 2022).

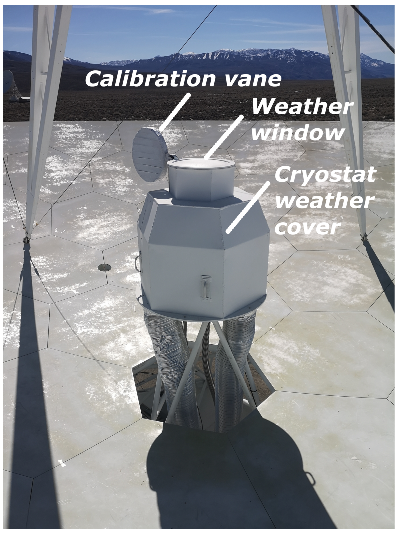
In fact, the COMAP-Pathfinder has already demonstrated two key milestones: the *Early Science (ES)* series of seven papers (of which Papers I and II of this thesis correspond to ES papers III and IV) and the *Season 2 Results* series of

Figure 12: COMAP-Pathfinder telescope from two different angles and times of day, with some humans in the foreground as a reference scale. Private photo.

three papers (where all are included in this work; Papers III-V). The ES papers presented the COMAP-Pathfinder instrument and showed that the systematics can be removed down to below the white noise level and that the data integrated down according to white noise expectations. As a result, we provide the first direct 3D LIM constraints of the cosmic CO(1–0) clustering (Paper II). The S2 results in Papers III–V continue on this path by significantly improving the ES data analysis, driving the upper limits on cosmic CO down another order of magnitude. We will return to the specifics of the data analysis of ES and S2 presented in the papers of this work in later sections.

The COMAP-Pathfinder fields a 10.4 m Cassegrain telescope, with a 1.1 m secondary mirror, (Leighton, 1977) at the Owens Valley Radio Observatory (OVRO), seen in Fig. 12. This corresponds to a full-width-half-maximum of the telescope beam of about $4.5'$ at 30 GHz. The telescope receiver has a 20-feed single-polarized detector focal plane array, one of which is blind and only used for diagnostics. Each detector has an independent electronics chain. To cool the detector array, the feeds themselves are located in a cryostat and kept at a system temperature of $30-60$ K.

The receiver is sensitive to frequencies in the $\mathrm{K_a}$-band, corresponding to 26–34 GHz. The 26–34 GHz signal is passed through a series of two down-converter modules (DCM1 and DCM2), located on the exterior of the receiver cryostat (internally, coined “saddlebags” for their visual appearance, see Fig. 13 right panel) and the telescope side cabin, respectively. The signal is then split into 4096 $\sim$ 2 MHz frequency channels by a ROACH2 spectrometer, distributed across two 4 GHz bands. Band A extends from $26-30$ GHz and band B from $30-34$ GHz. Bands A and B are divided into sidebands of width 2 GHz. The signal at all frequencies is boosted by the same low-noise amplifier (LNA) for a given feed to increase the signal-to-noise ratio (S/N) of the extragalactic CO. The instrument captures a spectrum every 20 ms and stores the resulting raw time-ordered data (TOD) on site before it is transferred to Oslo for data analysis. To calibrate the



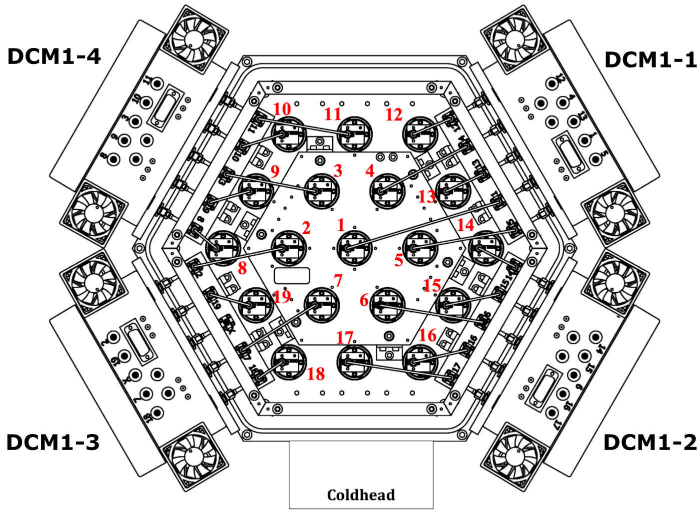


Figure 13: *Left*: COMAP-Pathfinder receiver array with the calibration vane, weather windows, and cryostat weather cover as fielded on the telescope. Private annotated photo. *Right*: Detector down-converter module (DCM1 or "Saddlebags") assignment inside the receiver backend. Courtesy of James Lamb.

data (this is shown later in Sec. 3.4), an ambient temperature-absorbing calibration vane, with temperature sensors, can be rotated in front of the detector array to create a reference load (see the left panel of Fig. 13). This overview of the instrument compiles some of the information presented by Cleary et al. (2022) and Lamb et al. (2022), and we refer the interested reader to their work for more details on the receiver.

3.1.2 Future COMAP phases

There are several planned future expansions to the COMAP-Pathfinder survey. The first of these is the recently funded COMAP-Wide, which will survey a 400 deg^2 area in the HETDEX spring field. It will overlap with the Low-Frequency Array AARTFAAC-12 (LOFAR A12) H I survey (van Haarlem et al., 2013) and the HETDEX galaxy survey (Gebhardt et al., 2021). Note that the field location, see Figs. 10 and 15, and the exact survey design are somewhat uncertain (and obtained through private communication) at the time of writing. During the four-year survey, two and a half years will be spent collecting around 5000 hours of observations using a duplicate of the 26–34 GHz COMAP-Pathfinder instrument. The primary goal of the survey is to place 30 times tighter constraints on the CO × HI power spectrum at $z \sim 7$ than would be possible with the state-of-the-art CO (COMAP; Papers III-V) and H I (HERA; (HERA Collaboration et al., 2023)) experiments individually in cross-correlation between the COMAP-Wide 400 deg^2 survey and the LOFAR A12 surveys. As the survey field will overlap with the HETDEX spring field, COMAP-Wide will also perform a cross-correlation between the HETDEX Lyα emitter (LAE) survey at $z \sim 3$.

To explore star formation at the EoR the COMAP-EoR and COMAP-ERA (extended reionization array) will map CO(1–0) at $z = 6$–8 using a 19-feed K$_u$-

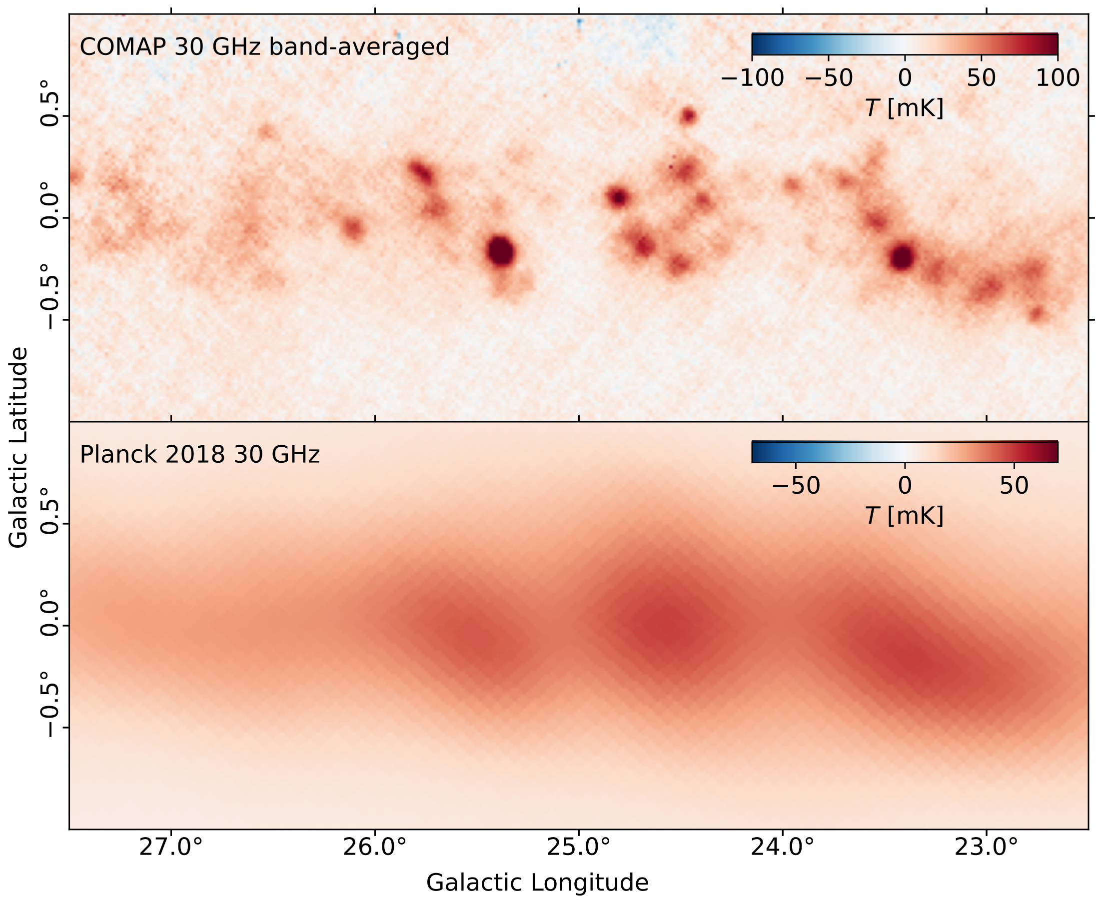


Figure 14: (*Top*) Section of the COMAP Galactic Plane Survey (Rennie et al., 2022) compared to the (*bottom*) Planck LFI 30 GHz map (Planck Collaboration II., 2020; Planck Collaboration IV., 2020) in the same patch.

band receiver on an 18 m new generation Very Large Array (ngVLA) prototype dish, as well as several copies of the COMAP-Pathfinder instrument (see previous subsection) sensitive to CO(2–1) from the same EoR redshifts (see Fig. 11). The observations in the K_a- and K_u-bands can then also be used to separate the CO(1–0) and CO(2–1) lines from the two redshift intervals. For more details on the future COMAP phases, we refer the interested reader to Cleary et al. (2022) and Breysse et al. (2022).

3.1.3 COMAP survey fields and scanning strategy

CO science fields and auxiliary Galactic plane survey

As described in Paper I, COMAP targets three main fields, Fields 1-3. For the field locations, see Figs. 10 and 15, are chosen to avoid bright point source foreground emission at 30 GHz (see backgrounds in Figs. 10 and 15 for Planck LFI 30 GHz map; Planck Collaboration I., 2020). To maximize observation efficiency, the fields are spread across right ascension. Additionally, to allow for cross-correlations with external galaxy catalogs (see, e.g., Chung et al., 2019, 2022a; Silva et al., 2021;

Breysse et al., 2019, for more details), the fields overlap (as seen in Fig. 15) with the Hobby-Eberly Telescope Dark Energy eXperiment (HETDEX) galaxy survey targeting LAEs at $z = 1.88 - 3.52$ (Gebhardt et al., 2021), the Spitzer/HETDEX Exploratory Large-Area (SHELA) survey area (Papovich et al., 2016), as well as the Sloan Digital Sky Survey (SDSS) Stripe 82 (Abazajian et al., 2009; Annis et al., 2014).

When the CO science fields are unavailable, COMAP conducts a Galactic plane survey (GPS) to measure the spectral energy distribution (SED) of anomalous microwave emission (AME; Rennie et al., 2022). The COMAP GPS has mapped the Galactic plane between $20^\circ < \ell < 40^\circ$ and $|b| < 1^\circ.5$ and conducted some detailed studies of, for instance, the λ-Orionis H II region (Rennie et al., 2022; Harper et al., 2025). Figure 14 shows part of the COMAP GPS of Rennie et al. (2022) together with the previous state-of-the-art 30 GHz GPS map by Planck LFI 30 GHz (Planck Collaboration II., 2020; Planck Collaboration IV., 2020). At 30 arcmin the Planck LFI 30 GHz beam is about 6.5 times as large as that of COMAP, giving COMAP a much higher spatial resolution as seen in Fig. 14. Additionally, due to COMAP's unique frequency resolution at 30 GHz, COMAP is ideally suited to perform high-resolution (both spatially and spectrally) surveys of the Galactic foregrounds and the region around the AME turnover in the continuum foreground SED. For examples of such measurements, we refer the interested reader to Rennie et al. (2022) and Harper et al. (2025).

Scanning strategy

When observing one of the main CO science fields (Fields 1–3 in Fig. 15), the data are divided into different observations, typically about one hour long, each associated with its own observation ID (obsID). During an observation, the telescope will point to the leading edge of the field and perform scans around a fixed azimuth (Az) and elevation (El) point while the field drifts across the sky. Subsequently, when the field has completely drifted past, the telescope is again repointed to the leading edge of the field, and a new scan is performed. Each observation is divided into several of these scans. This scanning strategy is illustrated in Fig. 16, both as seen from the horizontal telescope (Az-El) coordinates and as seen from equatorial coordinates where the field is stationary.

The scanning motion between repointings can be performed in several different ways, as shown in Fig. 17. The two panels of Fig. 17, respectively, show a Lissajous and constant elevation scanning (CES) type. The two scanning patterns can both be described as a harmonic motion of the form

$$\mathrm{az}(t) = A\sin(at + \phi) + \mathrm{az}_0 \tag{3.1}$$

$$\mathrm{el}(t) = B\sin(bt) + \mathrm{el}_0, \tag{3.2}$$

where the telescope pointing will oscillate around a fixed point $(\mathrm{az}, \mathrm{el}) = (\mathrm{az}_0, \mathrm{el}_0)$ with angular amplitude given by the parameters A and B. The phase shift between oscillations in azimuth and elevation is given by ϕ and the shape of the scanning trajectory is defined by a/b. Lissajous scans are characterized by two phase-shifted oscillations, while CES scans only move across azimuth (i.e., $B = 0$). A scan

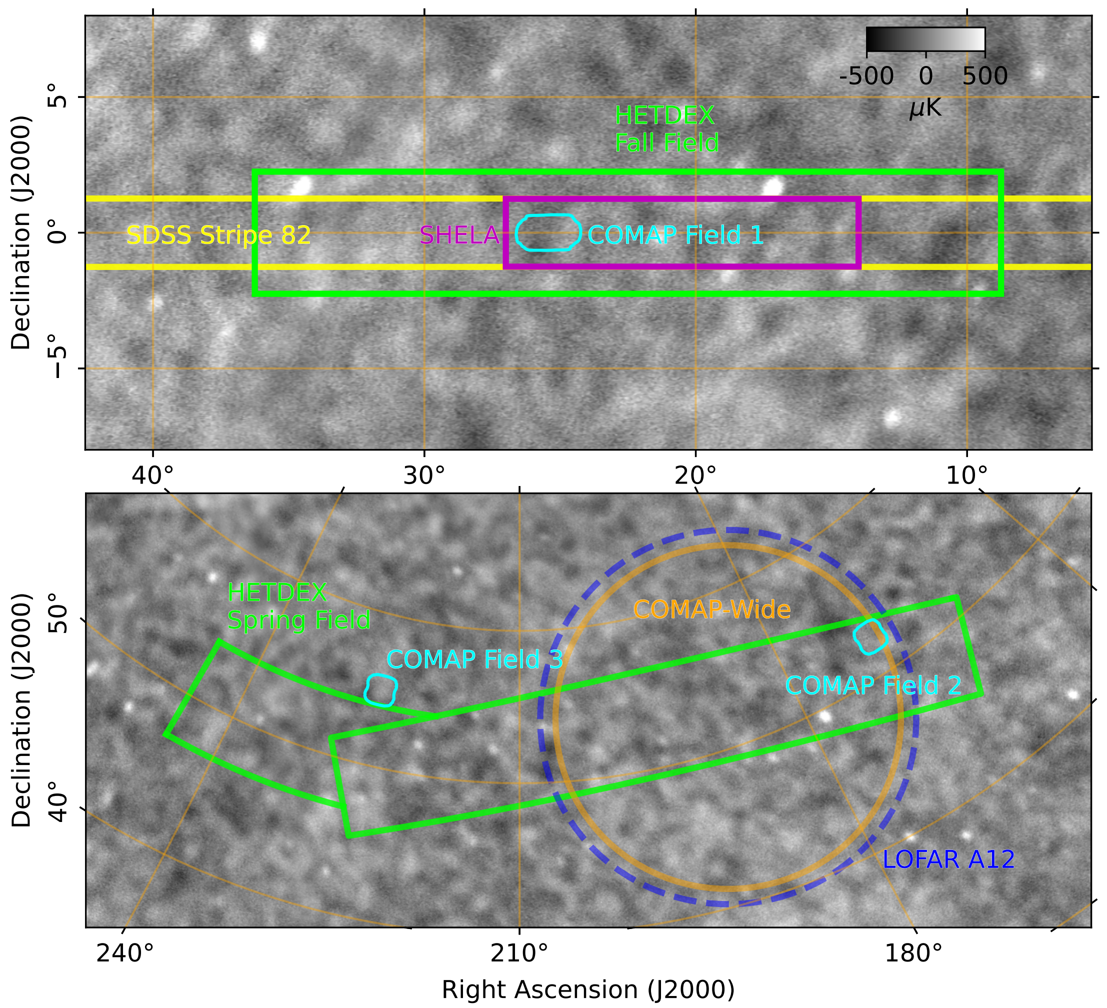


Figure 15: Zoom-in on the COMAP field locations seen in Fig. 10. The overlapping fields of HETDEX (Gebhardt et al., 2021), SHELA (Papovich et al., 2016) and SDSS Stripe 82 (Abazajian et al., 2009; Annis et al., 2014) are also shown for reference. Note that the COMAP-Wide and LOFAR A12 field position and extent are highly preliminary. The background shows the Planck LFI 30 GHz frequency map from the Planck Legacy Archive (Planck Collaboration I., 2020).

usually lasts between 3–10 min with an azimuthal scanning period of around 10–20 s (Paper I).

In the COMAP ES papers (Papers I and II), one of the main objectives was to test both scanning strategies as they each have their own advantages and disadvantages. Although generally superior in cross-linking and scanning speed, Lissajous is more prone to systematic errors, such as atmospheric and ground pickup due to elevation changes, as compared to CES scans (more on these systematics in Sec. 3.2). Indeed, one of the conclusions of the ES papers (Papers I and II) is that the Lissajous scans are challenging to clean properly, and thus, only CES scans are used in S2 of COMAP (Papers III–V).

Furthermore, in the ES papers elevations from about 30°–75° are used, while we limited our range to 35°–65° in elevation in the S2 papers because evidence for far-

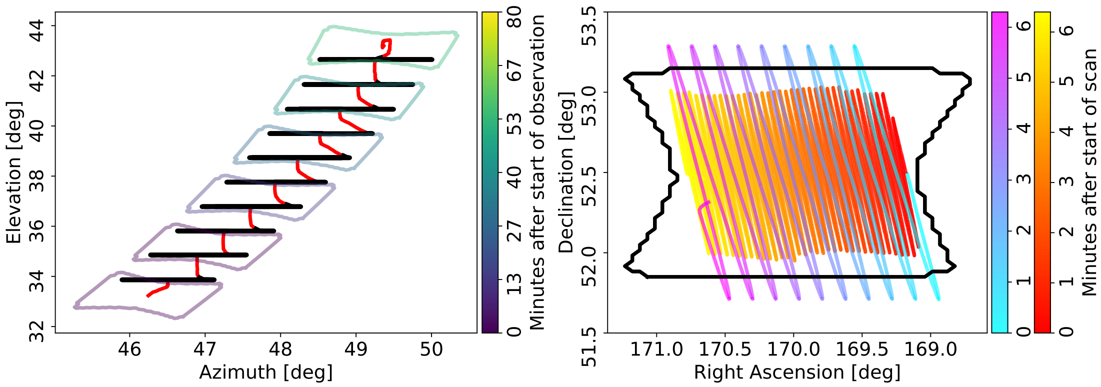


Figure 16: Example of COMAP scanning strategy as seen in local Az-El coordinates at OVRO (*left*) and in equatorial coordinates (*right*). In the left panel, hit contours of feed 1 are shown and are colored according to the field position in Az-El as time progresses in relation to the constant elevation scan pointing of the telescope (in black). The repointing between scans is colored in red. For readability, the field position is only shown for about every second scan. In the right panel, the hit contours of feed 1 are shown in relation to the pointing of two individual scans with two different azimuth speeds and are colored according to the progression of time.

sidelobe pickup at high and low elevations was found in the ES papers. Specifically, at high and low elevations, one or more sidelobes transition from hitting the sky to the mountains at OVRO. This causes sharp gradients in the signal that are difficult to remove in data analysis. Another change to the scanning strategy that happened about midway through the second season of observations was a reduction in the azimuth speed of the telescope drive due to wear in the mount drives. As seen in Fig. 16, showing examples of the two scanning strategies in equatorial coordinates, this resulted in the effective field footprint increasing somewhat and accumulating more hits at the edges of the field where the telescope now spends more time. We can also see how the telescope performs a lower number of sweeps in roughly the same amount of time.

3.2 The COMAP data model

As discussed in Sec. 3.1, the COMAP experiment aims to isolate the large-scale extragalactic CO signal in the data. To do so, we start by introducing the COMAP data model and illustrate how the data look like (Papers I and III), before continuing with an overview of the data analysis pipeline that aims to extract the CO emission in the next sections (Papers I–V). We will be somewhat brief when describing the low-level aspects of the data and the analysis thereof, as our main contributions to the COMAP analysis have been to higher levels of the analysis pipeline, such as mapmaking and power spectrum estimation.

The power output of the telescope at a given frequency channel ν, time t and detector i can, as explained in Paper I and III, be written as

$$P_{t\nu i} = k_{\mathrm{B}} \Delta\nu G_{t\nu i} T_{\mathrm{sys}}^{t\nu i}, \tag{3.3}$$

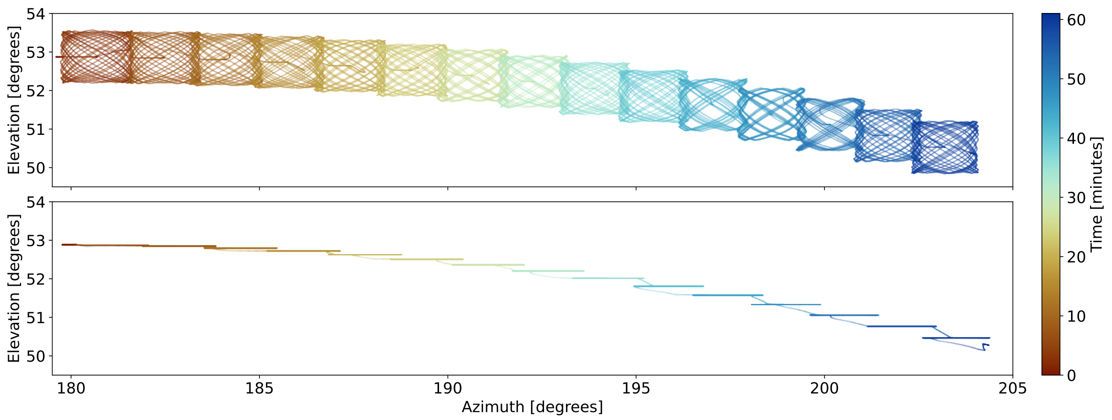


Figure 17: Example of telescope pointing in Az-El coordinates of two observations of Field 1, respectively, with a Lissajous (*top*) and constant elevation (*bottom*) scanning strategy. The two observations each have several scans, i.e., scanning periods between repointings. Courtesy of Jonas Lunde, see Paper I.

where the Boltzmann constant is given as k_{B}, the channel width as $\Delta\nu$, the time and frequency dependent gain by $G_{t\nu i}$ and the system temperature as $T^{t\nu i}_{\mathrm{sys}}$.

The system temperature, seen as a function of frequency in Fig. 18, quantifies all contributions to the signal and noise in the system and can be expanded into

$$\begin{aligned} T_{\mathrm{sys}} =& T_{\mathrm{receiver}} + T_{\mathrm{atmosphere}} + T_{\mathrm{ground}} + T_{\mathrm{CMB}} \\ &+ T_{\mathrm{foregrounds}} + T_{\mathrm{CO}}. \end{aligned} \tag{3.4}$$

The system temperature, which ranges from 30–70 K with a mean of around 44 K (see appendix Paper I for distribution), is thus a sum of the effective receiver noise temperature T_{receiver}, the contribution to the signal from the atmosphere $T_{\mathrm{atmosphere}}$, the ground pickup T_{ground}, foreground continuum emission $T_{\mathrm{foreground}}$, the CMB temperature monopole T_{CMB}, as well as the extragalactic CO emission we want to detect T_{CO}.

We can write Eq. (3.3) on a more explicit form as

$$\begin{aligned} d_{t\nu i} = \langle d_{t\nu i}\rangle(1+\delta g_{ti})\Big[1 &+ \mathsf{P}^{\mathrm{cel}}_{ti}\mathsf{B}_{\nu i}\left(\Delta s_{\mathrm{cont}} + \Delta s^{\mathrm{CO}}_{\nu}\right) \\ &+ \mathsf{P}^{\mathrm{tel}}_{ti}\mathsf{B}_{\nu i}\Delta s_{\mathrm{ground}} + s^{\mathrm{SW}}_{t\nu i} + n^{\mathrm{w}}_{t\nu i} + n^{\mathrm{corr}}_{ti}\Big], \end{aligned} \tag{3.5}$$

where the time averaged-data $\langle d_{t\nu i}\rangle$ is a normalization factor corresponding approximately to the product of the time-averaged gain system temperatures, $g_{\nu i} \equiv \langle G_{t\nu i}\rangle$ and $T_{\nu i} \equiv \langle T^{\mathrm{sys}}_{t\nu i}\rangle$ (absorbing the constants of Eq. (3.3) into $g_{\nu i}$). The shape of this term is illustrated in Fig. 19. In other words, the full gain $G_{t\nu i}$ is decomposed into a frequency and time-dependent part, $g_{\nu i}$ and $(1+\delta g_{ti})$ respectively. The gain fluctuations δg_{ti} are common for all frequencies at any given time and feed as the same LNAs process all channels. The sky signals from foregrounds and extragalactic CO, Δs_{cont} and Δs_{CO}, are projected from celestial coordinates into the time stream by the pointing matrix $\mathsf{P}^{\mathrm{cel}}_{ti}$ (see Sec. 3.1.3 for pointing) after being smoothed by the beam $\mathsf{B}_{\nu i}$. The beam-convolved ground

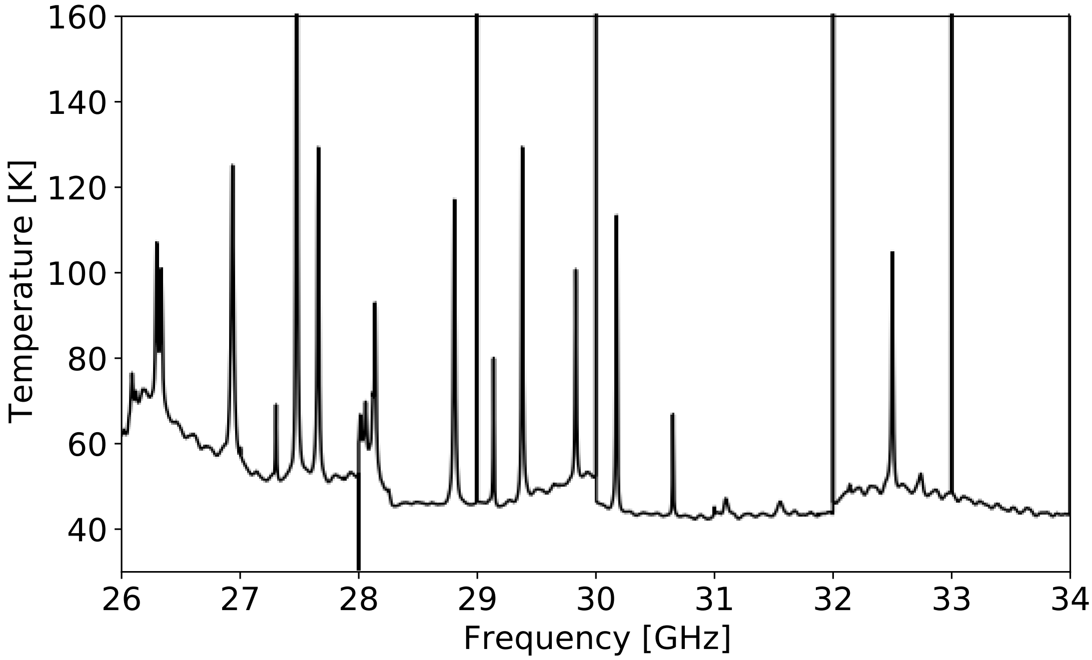


Figure 18: Typical time-averaged system temperature of Feed 1 over the experimental frequency bandpass. Courtesy of Håvard Ihle, see Paper I.

pickup map $\mathsf{B}_{\nu i}\Delta s_{\mathrm{ground}}$ (e.g., Fig. 22) is mapped into the time domain by the horizontal coordinate pointing matrix $\mathsf{P}^{\mathrm{tel}}_{ti}$. Note that the beam operator $\mathsf{B}_{\nu i}$ (see Fig. 20 for illustration) is absorbed into the signal sky and ground signal maps $\Delta s^{\mathrm{CO}}_{\nu}$, Δs_{cont} and $\Delta s_{\mathrm{ground}}$ in Paper I, but we explicitly include it here for clarity. We let $s^{\mathrm{SW}}_{t\nu i}$ denote time-, frequency- and feed-dependent standing waves (SW) in the data. Lastly, the terms $n^{\mathrm{w}}_{t\nu}$ and n^{corr}_{t} describe respectively the white and correlated noise. The instrumental noise is drawn from a zero-centered Gaussian distribution

$$n_{t\nu i} \sim \mathcal{N}\left(0, \frac{T_{\nu i}}{\sqrt{\Delta\nu\Delta t}}\right), \tag{3.6}$$

with a standard deviation following the radiometer equation, $\sigma^{\mathrm{w}}_{t\nu} = \frac{T_{\nu i}}{\sqrt{\Delta\nu\Delta t}}$. As we can see, the noise level of $\sigma_{\nu i}$ is determined by the integration time Δt, which means that the noise level decreases as more data are collected and combined. The correlated temperature noise $n^{\mathrm{corr}}_{t\nu i}$ of the COMAP telescope is mostly caused by atmospheric fluctuations and time-dependent standing waves, with a $1/f$ spectrum

$$P(f) = \sigma^{\mathrm{w}}_{t\nu}\left[1 + \left(\frac{f}{f_{\mathrm{knee}}}\right)^{\alpha}\right], \tag{3.7}$$

where the knee frequency f_{knee} and slope α determine the shape of the $1/f$ component. However, note that the gain fluctuations δg_{ti} are an additional and dominant source of $1/f$ noise.

In the following, we will summarize each of the temperature contributions in $\delta T_{t\nu i}$, contributing to the data model in Eq. (3.5), in terms of their brightness and stability.

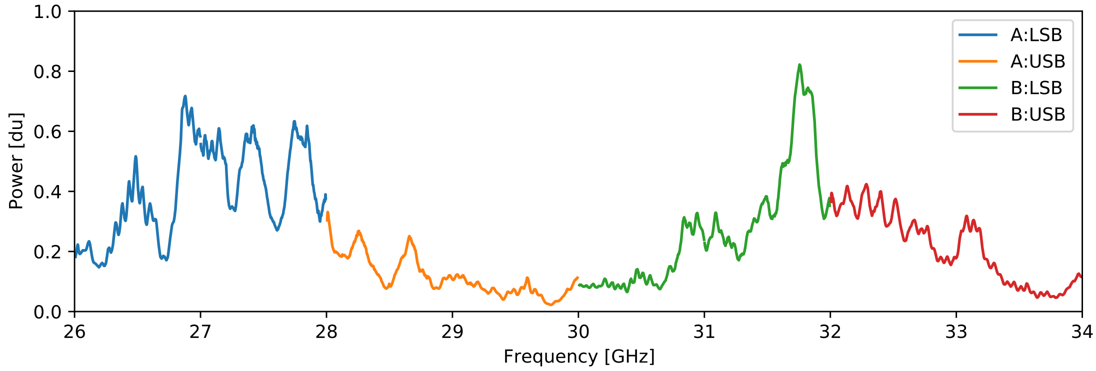


Figure 19: Typical time-averaged unfiltered and uncalibrated raw data per frequency for some detector of the COMAP telescope. The coloring marks the four different 2 GHz sidebands of the data. Courtesy of Jonas Lunde, see Paper I.

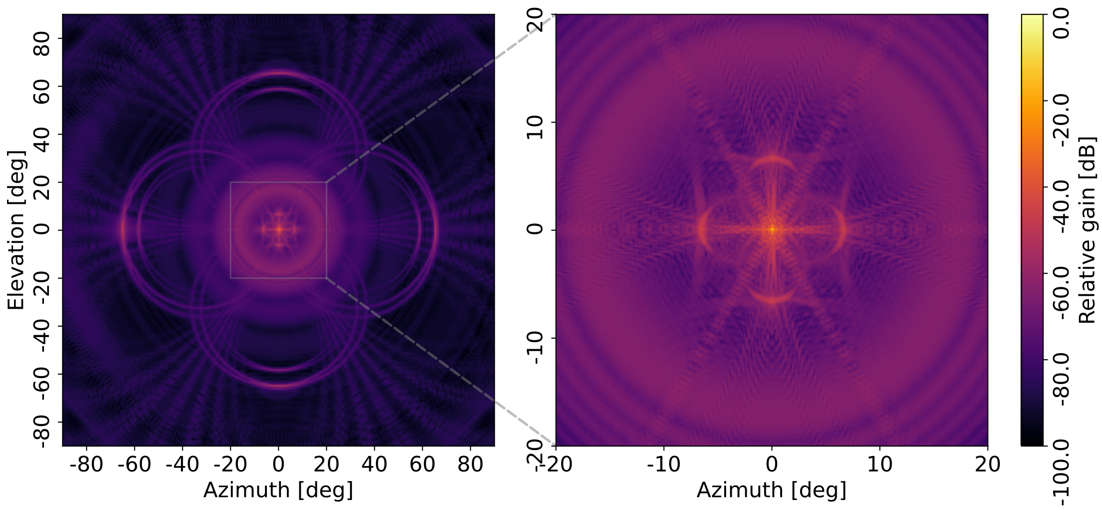


Figure 20: Simulation of COMAP beam in units of decibel (dB) at 30 GHz made with TICRA Tools by Lamb et al. (2022). The right panel shows an inset of the inner $20° \times 20°$ of the beam in the left panel.

3.2.1 Receiver noise

The receiver temperature is the largest contribution to the noise in the system due to the finite temperature of the cryostat that causes thermal noise. It contributes about 10–30 K to the system temperature and is very stable due to the HEMT LNAs used by COMAP (Lamb et al., 2022). In the left panel of Fig. 21, an example is shown in which the raw data are dominated by instrumental noise.

3.2.2 Atmospheric signal

The atmosphere's temperature contributes an overall 15–20 K to the system temperature. The atmospheric signal is highly dependent on the local environment at OVRO and can vary with weather and pointing (the latter of which we will return to in Sec. 3.4.2). It is correlated on time scales of several seconds but behaves

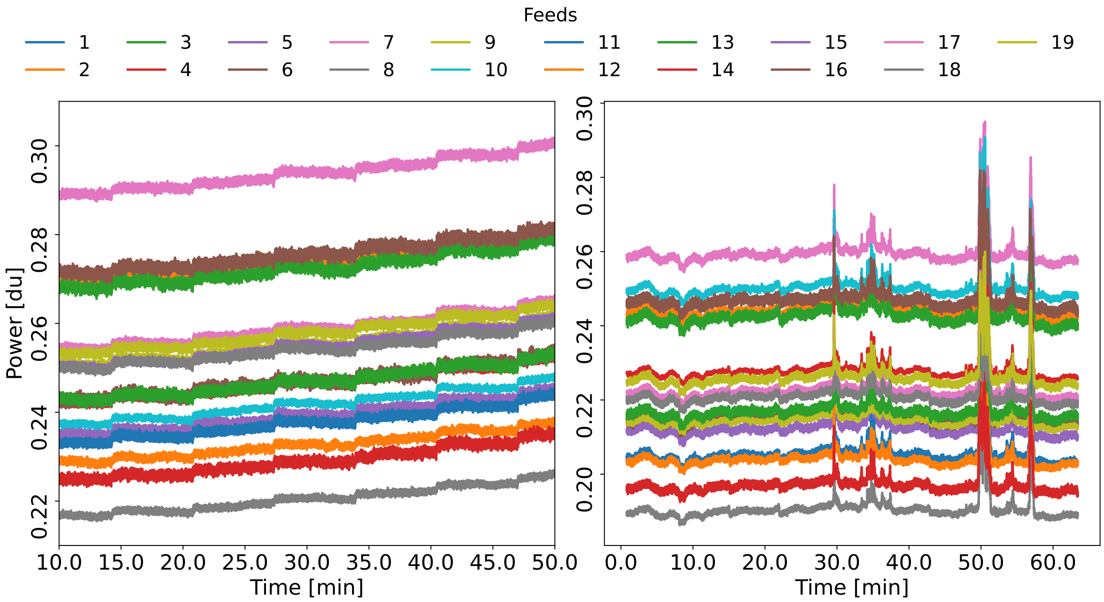


Figure 21: Examples of raw data time-streams taken during different observing conditions for each feed and averaged over frequencies. The observation used in the left panel is taken under good conditions and dominated by instrumental noise. Each “step” in the “staircase” represents a constant elevation scan between repointings. In the right panel, an observation is shown where there are clear signs of bad weather contamination correlating across feeds.

as uncorrelated noise on time scales of hours or longer. Atmospheric temperature fluctuations represent one of the sources of $1/f$ noise in the system. Due to different feeds looking through the same atmosphere, the atmosphere pickup is highly correlated across feeds and frequencies, as can be seen in the right panel of Fig. 21, showing some strong weather spikes in all feeds.

3.2.3 Ground pickup signal

Radiation from the ground will be picked up by the telescope as spillover that diffracts around the secondary mirror, as well as from reflections off the secondary mirror support legs, hitting the detector array. The latter is highly pointing dependent and causes far-sidelobe pickup from the landscape at OVRO as well as potentially the sun and moon at certain times. A TICRA Tools[1] simulation, made by Lamb et al. (2022), is seen in Fig. 20. It shows the telescope beam with its $\sim 65^\circ$ far-sidelobes at 30 GHz. The far-sidelobe contribution can cause sharp transitions in the signal pickup, which can be challenging to remove in the data analysis. Therefore, certain elevation ranges and proximity to the Sun and Moon are avoided (see earlier Sec. 3.1.3). In Fig. 22, the OVRO ground profile, convolved with the COMAP beam model, shows the approximate level of ground pickup expected, as well as the sharp gradients seen in certain elevation ranges. Overall, the ground signal makes up about 5–6K of the total system temperature.

[1] See https://www.ticra.com/software/grasp/

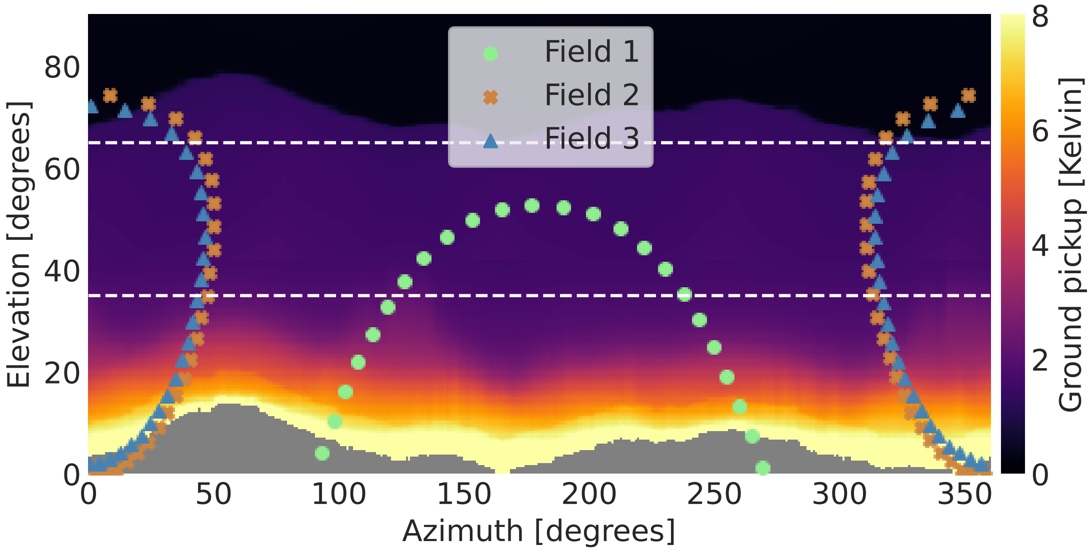


Figure 22: Ground profile at OVRO convolved with the COMAP beam model, showing the approximate expected scale of the ground pickup signal. The three COMAP field trajectories are overplotted, in addition to the elevation range used for CO-science in COMAP S2. Courtesy of Jonas Lunde, see Paper III.

3.2.4 Continuum emission

The telescope will pick up radiation from continuum sources such as synchrotron, free-free, thermal, and spinning dust emissions from the Galactic foreground. These are static in time but not isotropic on the sky and make up about 1 mK of the total signal. The CMB, which, due to its cosmological nature, is static and isotropic, contributes 2.7 K to the system temperature. Lastly, there can be some extragalactic continuum foregrounds, which we, however, treat on the same footing as the Galactic foregrounds. The level of continuum emission within and around our fields can be seen in Figs. 15 and 10, showing the Planck LFI 30 GHz map (Planck Collaboration I., 2020) around our Field locations.

The Galactic foregrounds and CMB signals have in common that they are spectrally smooth and behave approximately (to first order) as a flat monopole in our degree-scale COMAP fields in the angular direction. This is in sharp contrast to extragalactic CO, which rapidly varies in frequency and, as we will see later in Sec. 3.4 makes it easier to separate the two.

3.2.5 Standing waves

Reflections in the telescope optics or between the ends of cables in the electronics give rise to standing electromagnetic waves in the COMAP data. Because most of these cavity lengths will be constant in time, so will the standing wave signal. This can then be easily subtracted by removing a static signal. However, time dependence in the cavity length, e.g., from mechanical vibrations or stretched cables (which can be different from detector chain to detector chain), can lead to non-trivial standing wave contamination of the data. Indeed, as we will discuss in

Sec. 3.4, this is one of the most challenging systematics to handle in the COMAP data analysis as standing waves can be highly correlated with both frequency and pointing at the same time. As such, they behave similarly to a static sky signal. The standing wave signal usually has an amplitude of around 10–300 mK.

3.2.6 Extragalactic CO

The targeted extragalactic CO emission is highly subdominant in amplitude to the other signal contribution with only the order of a few μK. This illustrates why LIM (in general and) with CO is a dynamic range problem that tries to isolate fluctuations of the order $\lesssim 10^{-6}$ compared to the instrument noise, systematics, and foregrounds. As mentioned, the CO emission fluctuations are expected to vary rapidly along the line of sight and can thus be separated from the continuum emission. Additionally, as a cosmic background, the CO emission is static and isotropic in the sky.

3.3 COMAP seasons and differences between them

As mentioned earlier, the work done for this thesis covers two rounds of COMAP publications; Early Science (or S1), and S2. We summarize the duration and observation time in each season, as well as some key notes, in Table 1. In sum, the raw data volume in S2 is about 3.4 times that of Season 1.

However, not only did the amount of observational time in each season change, but also several other aspects of the observational strategy, instrumental parameters, and the pipeline. Some of these changes are motivated by lessons learned during the Early Science analysis, while others were necessitated by new discoveries or events. Although more details on the analysis pipeline and data selection will follow in later sections (3.4 and 3.5), the following overview will condense some of the largest points of change throughout the analysis mentioned in Papers I–IV:

Observing

- To avoid strong gradients in the ground pickup from far-sidelobes seen in Fig. 22, the allowed elevations at which to observe were changed from $30° - 75°$ in ES (i.e., Papers I and II) to $35° - 65°$ in S2 (i.e., Papers III-V; see Fig. 22). This was done because the higher and lower elevations showed hints of residual sidelobe pickup in the ES data, resulting in these data being removed from the ES analysis.

- Because Lissajous scans showed evidence for residual systematic effects after cleaning the data from changes in elevation, one of the key conclusions from Papers I and II is to abandon Lissajous scans in favor of only using only constant elevation scans.

- About halfway through the second season of observations, the maximum allowed scanning speed of the azimuth motors was halved because of signs of wear and tear. The fast- and slow-moving parts of S2 are referred to as Season 2a and Season 2b in Paper III; see Table 1).

- About the same time as the drive speed reduction, the spectrometer clock frequency was changed from 4.000 GHz to 4.250 GHz to widen Bands A and B slightly by raising the spectrometer's Nyquist frequency. By widening the band's frequency samples with high aliasing power fall outside the bandpass of interest for COMAP (26–30 GHz for Band A and 30–34 GHz for Band B), resulting in a lower number of discarded frequency channels from aliasing cuts within the 26–30 GHz bandpass. Thus, data captured before and after S2b are spaced on, respectively, a 1.953 MHz and 2.075 MHz channelization.

Pipeline

- The analysis pipeline used for the S2 release (Papers III-IV) is for the most part a Python/C++ reimplementation of the Early Science (Papers I and II) FORTRAN pipeline, although it remains algorithmically very similar. This was done to increase development speed, maintainability, and optimize pipeline runtime performance.

- Several steps in the pipeline were either added or changed to mitigate newly discovered systematics or improve existing algorithms, as will be explained further in the next sections. Among these are the additions of a per-feed PCA filter, a map-PCA filter, and some modifications to the power spectrum methodology and null tests between the two publication series.

- Data selection, which will be described more in Sec. 3.5, was changed significantly between S1 and S2, tightening some and loosening other data cuts, and results in an increase from 6.8% to 21.6% of the theoretical maximum achievable power spectrum sensitivity.

3.4 The low-level analysis pipeline

Now that we have introduced the COMAP data model and highlighted some major changes between the COMAP-Pathfinder seasons, we continue introducing the data analysis pipeline. The main goal of the analysis pipeline is to clean and calibrate the data, leaving only noise and extragalactic CO emission, to construct the world's first large-scale CO-LIM map from which we can constrain the physical properties of the star-forming Universe.

Note that the current iteration of the COMAP analysis pipeline is a so-called classical linear pipeline, which takes in data, filters it, and provides some final cleaned data product. This is a good first approach while trying to understand the data. However, there are more well-motivated Bayesian analysis frameworks, though these usually also become computationally more challenging with large datasets like that of COMAP. Nevertheless, we will later in Ch. 4 consider the `Commander3` framework as a possible future improvement of the COMAP pipeline.

Season	Time span	Observing time [h]	Notes
Season 1	05/2019 – 08/2020	$5,200$	Employed 50/50 Lissajous and CES type scans. Before azimuth slow-down (Fast Az).
Season 2a	11/2020 – 05/2022	$7,900$	Only CES type scanning. Before azimuth slow-down (Fast Az).
Season 2b	05/2022 – 11/2023	$4,400$	After azimuth slow-down (Slow Az) and spectrometer sample frequency change.

Table 1: Observational seasons of the COMAP-Pathfinder, showing the period over which each season was observed and how much observation time they contain. For more details on changes between the seasons see Sec. 3.3. The table is adapted from Paper III, courtesy of Jonas Lunde. Note the terminology of Paper IV refers to Seasons 1 and 2a together as "*Fast Azimuth*", and Season 2b as the "*Slow Azimuth*" data.

In the following, we will go through the steps in the low-level analysis pipeline, as schematically drawn in Fig. 23, and describe how the filters affect the data as we go. The details of the COMAP low-level pipeline are found in Papers I and III, respectively, for the COMAP ES and S2 releases.

3.4.1 Level 1 data and data segmentation

The raw data from the telescope are stored in the so-called *level 1* format at OVRO before being sent over to Oslo for data analysis. It contains both the raw time-ordered data (TOD) for all frequencies and feeds and needed housekeeping data for all scans captured during one obsID. The very first step of the pipeline is to iterate over all obsIDs in the level 1 files in the `scan_detect` code, split up all obsIDs into their constituent scans, and build a database over all observations. The observation database and level 1 files are subsequently processed by the level 2 generator (`l2gen`) code.

3.4.2 Level 2 data – time-domain filtering and calibration

The level 2 files, which are produced by the `l2gen` code for each 3–10 minutes scan, contain the cleaned and calibrated TODs for each feed and frequency, as well as all housekeeping data for later data selection. In the following, we will summarize each step in `l2gen`, which are also schematically illustrated in a flow-chart in Fig. 23, and we refer the interested reader to Papers I and III for more details than are mentioned in the following.

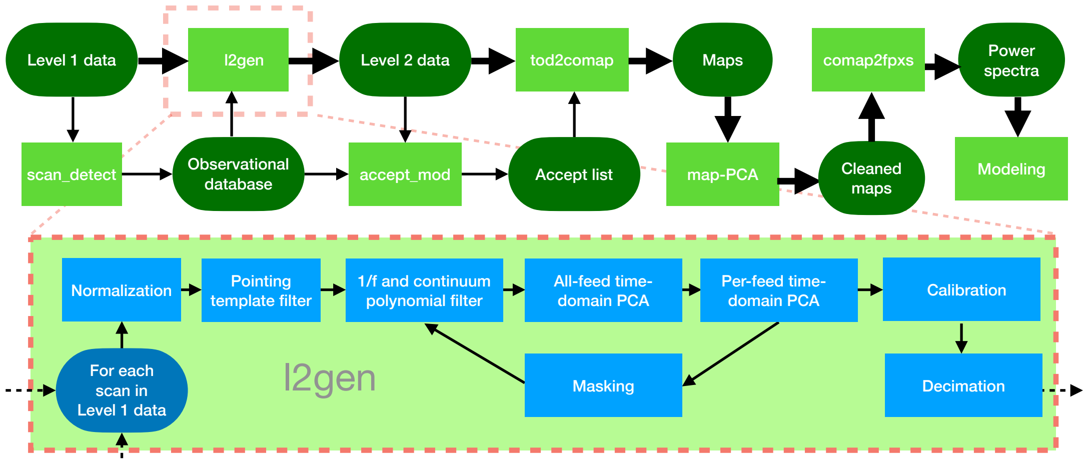


Figure 23: Flow-chart showing the COMAP pipeline. The chart is a compromise between the ones shown in respectively Papers I and III, showing all steps (dark ellipses) in both the ES and S2 pipelines and how the data products (light squares) flow through each pipeline step. The "flow" of the main data products is indicated by thick arrows, while thin arrows indicate auxiliary data flow. As the `l2gen` step in itself contains numerous individual steps, a flow chart of the filtering steps is included as an inset. For a full explanation of each step, see the main text of Sec. 3.4.

Normalization

As discussed in Sec. 3.2, the raw data are dominated by the experimental bandpass. However, this bandpass shape is simply a normalization constant that causes the noise properties of the raw data in each channel and feed to be different. We need to flatten out the bandpass seen in Fig. 19, so each channel has the same noise properties before any further pipeline steps are applied.

To do so we divide the raw data by the mean and subtract one so that

$$d_{t\nu i}^{\mathrm{norm}} = \frac{d_{t\nu i}}{\langle d_{t\nu i}\rangle} - 1, \tag{3.8}$$

where the normalized TOD $d_{t\nu i}^{\mathrm{norm}}$ now fluctuates around zero with the same noise level everywhere. This is illustrated in the transition between the left and right panels of Fig. 24. The data are dominated by the frequency-dependent gain fluctuations prior to normalization, and the (frequency-)common mode $1/f$ gain fluctuations are visible after filtering.

Specifically, we want to allow $\langle d_{t\nu i}\rangle$ to drift slowly in time as there could be slowly varying modes in the temperature-gain product $\langle G_{t\nu i}T_{t\nu i}^{\mathrm{sys}}\rangle$. The slowly varying temperature-gain fluctuations over the bandpass are found by computing the running mean of the raw data:

$$\langle d_{t\nu i}\rangle = \mathcal{F}^{-1}\{\mathcal{F}\{d_{t\nu i}\}W\}, \tag{3.9}$$

where $\mathcal{F}$ denotes Fourier transforms and the Butterworth filter kernel W is given by

$$W = \left(1 + \left(\frac{f}{f_{\mathrm{knee}}}\right)^{-\alpha}\right)^{-1}. \tag{3.10}$$

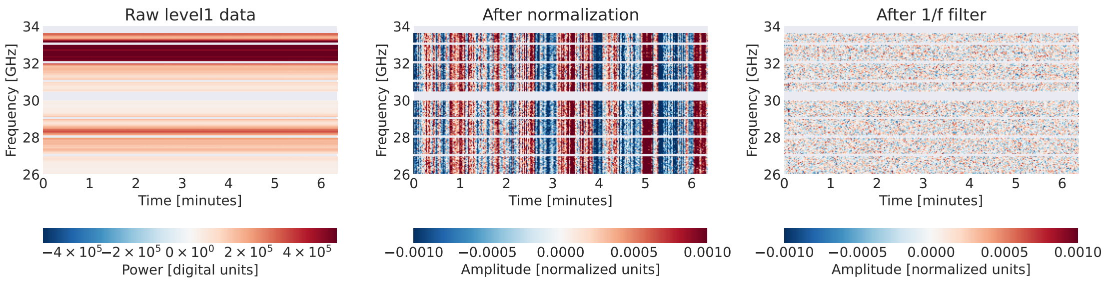


Figure 24: Effect of normalization and $1/f$ filtering for a single scan and feed. The raw TOD as a time-frequency waterfall (*left*) is dominated by the gain variations of the experimental bandpass in frequency (see Fig. 19). The normalized data (*middle*) is dominated by $1/f$ gain fluctuations that are highly correlated across frequency channels. After the polyfilter (*right*), the $1/f$ fluctuations are removed, and the data are dominated by white noise. Courtesy of Jonas Lunde, see Paper III.

The slope and knee frequency parameters are set, respectively, to $\alpha = 4$ and $f_{\text{knee}} = 0.01\,\text{Hz}$ to remove all temporal modes longer than about 100 s. In Fig. 25, an example is shown that more clearly illustrates the removal of slowly varying temporal modes in the TOD for a given frequency channel and feed. Thus, this filtering step also removes any monopoles of the COMAP field. After filtering, the data $d^{\text{norm}}_{t\nu i}$ have the same noise level on all frequency channels.

Pointing template removal

After normalizing the experimental bandpass, we can remove the signal that is correlated to the telescope pointing in horizontal coordinates at OVRO. In general, we can model the pointing template as

$$d^{\text{pointing}}_{t\nu i} = \frac{\gamma_{\nu i}}{\sin \text{el}_{ti}} + \alpha_{\nu i}\text{az}_{ti} + c, \tag{3.11}$$

where azimuth and elevation at any given time t, and detector i, are denoted as az_{ti} and el_{ti}. The free parameters $\gamma_{\nu i}$ and $\alpha_{\nu i}$ are fit for by χ^2 minimization separately per frequency and feed. The parameter c denotes a constant offset, which can be removed by subtracting the average of the template. Note that the first term of Eq. (3.11) corresponds to the optical depth $\tau(\text{el}) = \tau_0/\sin(\text{el})$, where we assume that the atmosphere can be modeled as a flat slab with optical depth at zenith given by τ_0. The second term, meanwhile, is linear in azimuth and can pick up azimuth-correlated signals, such as the sidelobe signal from the ground. As such, only the azimuthal template is subtracted when performing a CES scan because there are no changes in elevation, that is, $\gamma = 0$. Meanwhile, in a Lissajous scan, both azimuth and elevation templates are subtracted. Therefore, since the second season of COMAP (Papers III–V) only employs CES scans, only the azimuth template is applied.

By subtracting the pointing template from the normalized data:

$$d^{\text{az/el}}_{t\nu i} = d^{\text{norm}}_{t,\nu i} - \frac{\gamma_{\nu i}}{\sin \text{el}_{ti}} - \alpha_{\nu i}\text{az}_{ti} - \left\langle \frac{\gamma_{\nu i}}{\sin \text{el}_{ti}} - \alpha_{\nu i}\text{az}_{ti} \right\rangle, \tag{3.12}$$

the horizontal coordinate pointing correlations are removed. The average of the template is subtracted to explicitly ensure a zero mean of the data (although note that this is only explicitly done in Papers I and II). In the second-season publications (Papers III-V) this filter was performed independently for west- and east-moving sweeps, while the Early Science version of the filter (Papers I and II) performed this fit jointly. This change was included to handle directionally dependent systematic effects observed in the S2 data.

The effect of this filter on a Lissajous scan is shown in the second row of Fig. 25, where we clearly can see the sinusoidal atmospheric contribution of the atmosphere due to elevation changes being removed by the pointing template.

Polynomial filter

Now that the effects of the experimental bandpass and pointing correlated atmospheric correlations are taken into account, we can consider the contributions to the correlated noise and continuum emission in the data. As mentioned earlier in Sec. 3.2, the $1/f$ correlated noise fluctuations are shared across frequency channels in a given feed and sideband due to common LNAs and a highly feed-frequency correlated atmospheric fluctuation pickup. This is illustrated in the middle panel of Fig. 24.

Additionally, the continuum emission present in the data, from Galactic foregrounds or the CMB, all behaves approximately as a smooth low-ordered polynomial in frequency. Meanwhile, the targeted fluctuations of extragalactic CO are expected to be sharply peaked at certain frequencies corresponding to density fluctuations along the line-of-sight.

Therefore, we can remove both the correlated $1/f$ noise and continuum emission in our data by simply fitting and removing a low-order polynomial

$$d_{t\nu i}^{\mathrm{poly}} = d_{t\nu i}^{\mathrm{az/el}} - (c_{t,0} + c_{t,1}\nu + c_{t,2}\nu^2 + \cdots) \tag{3.13}$$

from the data in every time-step t. The free parameters c_i are independently fit for each time step t, and for each sideband and detector (with shared electronics). Although this filter is fully linear, it constitutes one of the most aggressive filters of the low-level analysis pipeline because of the large number of free parameters. In both ES and S2 publications (Papers I and III) a first-order polynomial is used.

In the right panel of Fig. 24 we see an example of the data after removing the correlated noise fluctuations using this polynomial (polyfilter). In addition, in Fig. 25 one can see an example of how the polyfilter also removes some of the noise at a frequency slice of the data.

Time-domain principal component analysis

The final filter that is applied in the time-domain is the principal component analysis (PCA; first developed by Pearson, 1901) filter. The purpose of the PCA filter is to remove residual systematic effects that the other filters have not managed to remove and that correlate across multiple frequencies and feeds.

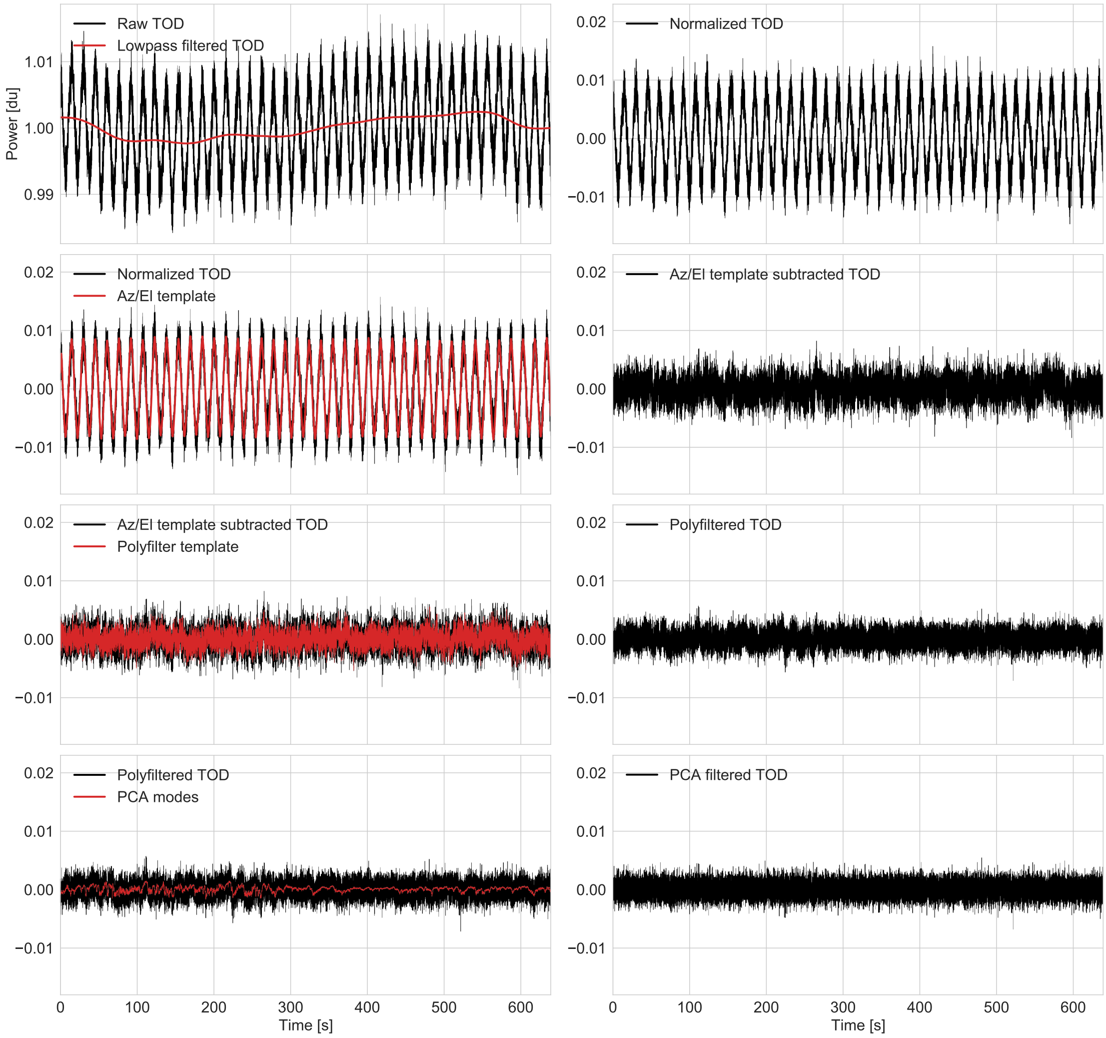


Figure 25: Time-ordered data at some given frequency and feed at different stages of filtering. The two columns, respectively, show the data before and after each filtering step in `l2gen` (one per row), where the red graph in the left column represents the components to be subtracted from the data. Courtesy of Jonas Lunde, see Paper I.

Moreover, there are several ways to think about a PCA, some more general than others. To start with, we introduce the PCA filter as we describe it in Paper I. The TOD over all frequencies can be written in matrix form as

$$\mathsf{D} = \begin{bmatrix} d_{11} & \cdots & d_{1n_{\text{time}}} \\ \vdots & \ddots & \vdots \\ d_{n_{\text{freq-feed}}1} & \cdots & d_{n_{\text{freq-feed}}n_{\text{time}}} \end{bmatrix}, \tag{3.14}$$

where we let n_{time} denote the number of time samples in a scan, and $n_{\text{freq-feed}}$ is either the number of frequency channels or frequency channels and feeds, depending on whether to perform the analysis on all feeds jointly or individual feeds. The corresponding principal components of the data $d_{t\nu}$ can then be defined as the eigenvectors w_t^k of the covariance matrix $\mathsf{C} = \mathsf{D}^T\mathsf{D}$. The basis of eigenvectors is ordered by the corresponding eigenvalues λ_k, such that the eigenvector with the highest eigenvalue describes the direction in data space of the highest explained variance.

To find the corresponding frequency vector a_ν^k we can project the eigenvector back onto the data for each frequency and feed

$$a_\nu^k = \sum_{t=1}^{n_{\text{time}}} d_{t\nu} w_t^k, \tag{3.15}$$

where $d_{t\nu}$ denotes the data after the previously described filter steps.[2] To filter the data we can then remove the outer product of the time and frequency vectors for each principal component k:

$$d_{t\nu}^{\text{PCA}} = d_{t\nu}^{\text{poly}} - \sum_{k}^{n_{\text{comp}}} w_t^k a_\nu^k, \tag{3.16}$$

with the number of removed principal components n_{comp}. In Paper I we subtract a static number $n_{\text{comp}} = 4$, while in Paper III a dynamic number of components is subtracted based on known approximations for a noise matrix.

Meanwhile, we can also define the PCA problem as done in Paper III where we want to find principal components w_t^k and a_ν^k such that the quantity

$$\sum_{t,\nu} \left(d_{t\nu} - \sum_{k=1}^{n_{\text{comp}}} w_t^k a_\nu^k \right)^2 \tag{3.17}$$

is minimized. Although completely equivalent to the earlier notation in this section, this way of formulating the problem allows for an easier generalization of the standard PCA when including noise weights on the data (see later Sec. 3.4.4.

The PCA will find a data reconstruction

$$d_{t\nu} \approx \sum_{k=1}^{n_{\text{comp}}} w_t^k a_\nu^k \tag{3.18}$$

that will perfectly describe the data if the entire set of principal components is used. Specifically, the principal components that explain most of the total variance of the data (largest eigenvalues λ_k) will also contain most of the information in the data. Principal component analysis is therefore often used to perform dimensionality reduction by removing the lowest principal components of a dataset that tend to contain only noise (see, e.g., Jia et al., 2022; Greenacre et al., 2022, for reviews). However, in the COMAP dataset we assume that systematic effects are orders of magnitude larger in amplitude than the extragalactic CO. Hence, the systematic effects will usually be well decomposed by the first couple of principal components while leaving the signal untouched. Furthermore, systematic effects like standing waves are usually well described by an outer product (seen for instance in the

[2] Equivalently, one can find the principal component basis of the data matrix by performing a singular value decomposition of the data; $\mathsf{D} = \mathsf{U\Sigma W}^T$. In that case, the data will be described by a basis of frequency and time vectors corresponding to the columns of U and W, as well as singular values σ_k along the diagonal of Σ corresponding to the square root of the eigenvalues λ_k.

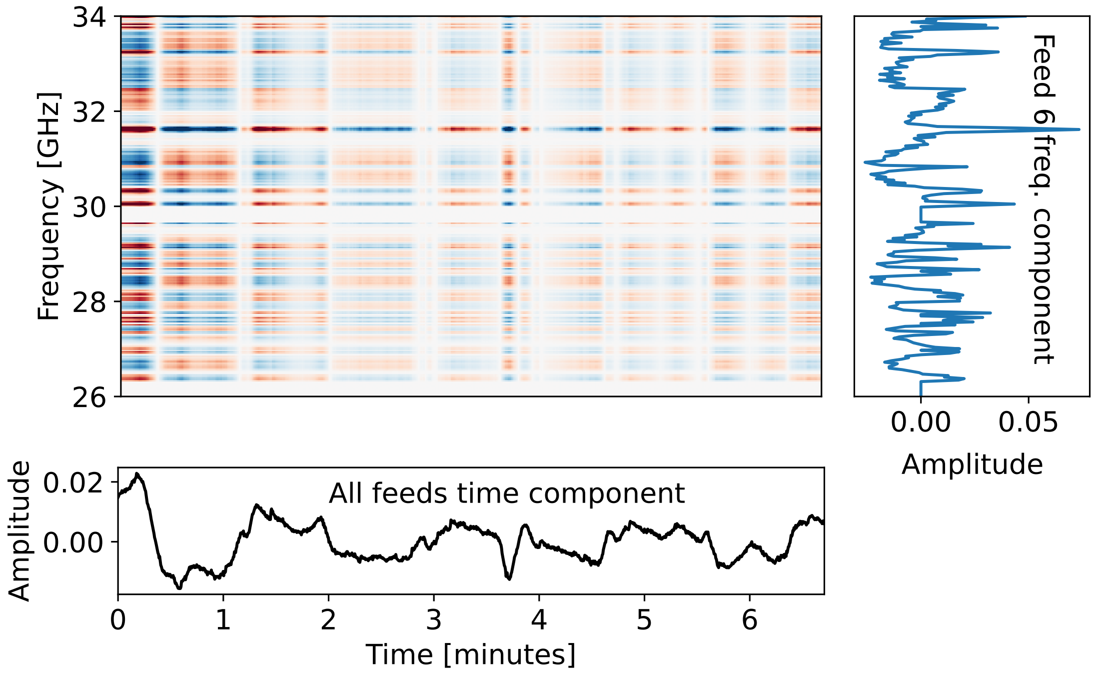


Figure 26: First principal component reconstruction of scan 3354205 using the all-feed PCA filter. The waterfall plot is the reconstruction of the TOD as captured by feed 6 and formed by the outer product of the time vector w_t^0 common to all feeds (*lower*) and the corresponding frequency vector of feed 6 a_ν^0 (*right*). The color range of the waterfall plot is $\pm 5 \cdot 10^{-4}$ in normalized TOD units. Courtesy of Jonas Lunde, see Paper III.

Taylor expansion of Eq. 2.2 by Chung, 2022b, around small electromagnetic cavity lengths) in time and frequency (or, as we shall see later in Sec. 3.4.4, celestial coordinates and frequency). Standing waves are, therefore, well modeled by only a few leading principal component vectors. The CO signal, meanwhile, is more complicated and requires more principal components to be modeled, which builds a further safeguard against signal loss.

As hinted earlier in this section, there are two versions of this PCA filter: the *all-feed* which is first included in the ES pipeline shown in Paper I, and the new *per-feed* PCA filter which is added to the S2 pipeline in Paper III. Both versions are algorithmically equivalent and can be described by the same equations as above. As their name indicates, the all-feed and per-feed PCA are, respectively, performed on a data matrix of all feeds jointly (shape $(n_\text{freq}n_\text{feed}, n_\text{time})$) and separately on the data matrices for each feed (shape $(n_\text{freq}, n_\text{time})$). For the per-feed PCA, we also found that down-sampling the data in frequency (letting $n_\text{freq} \to n_\text{freq}/16$ through standard inverse noise weighting) increases the effectiveness of the filter in identifying large-scale systematic modes. As shown in `l2gen` flow-chart of Fig. 23 the all-feed filter is performed before the per-feed PCA. Treating all feeds jointly will target feed-correlated systematics, like a common-mode residual atmosphere or standing waves in the shared optics. Meanwhile, the per-feed PCA is meant to target components that are specific to each feed, such as standing waves in the independent detector electronics chain. As such, the filters will only target CO

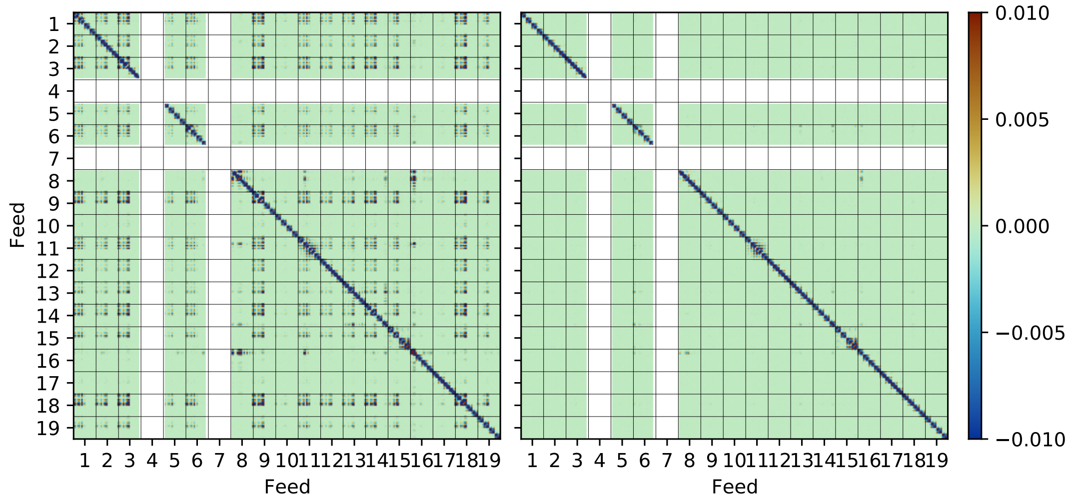


Figure 27: Example from Paper I showing the correlation between the feeds and frequencies of a typical scan before (*left*) and after (*right*) the all-feed PCA time domain filter. We note that feeds 4 and 7 are masked in this particular scan. Courtesy of Håvard Ihle, see Paper I.

signal modes that span the entire survey volume, mostly leaving it unaffected.

In Fig. 26, we show an example from Paper III of the first principal component reconstruction and the basis vectors that form it. Figure 27 shows a feed-frequency correlation matrix before and after subtracting the four leading modes of the all-feed PCA filter. This removes most of the correlations, and the result is left without significant off-diagonal correlation structures, indicating a successful removal of the common mode signal.

Masking

In the COMAP pipeline, there are two basic frequency masking procedures. The first mask removes frequency channels with known problems at the very beginning of the pipeline before any filters are applied. For example, we mask the edges of each frequency band due to aliasing, which can induce correlations of the order 10 %. Furthermore, in the ES pipeline (Paper I), channels with system temperature higher than 80 K were masked, while we in the S2 pipeline (Paper III) introduce a dynamical approach. Specifically, we remove channels with a system temperature greater than 5 K over its running median. Lastly, some channels are manually flagged if they are known to correlate with systematic effects.

Secondly, we mask channels that show signs of significant residual systematic correlations between frequency channels after applying the poly- and time-domain PCA filters to the data. Specifically, a χ^2 test is applied to boxes and stripes along the frequency-frequency correlation matrix. Boxes or stripes with correlations larger than 5σ above the expected correlation between Gaussian random variables and the expected channel-channel correlations within each sideband induced by the polyfilter are masked out. Additionally, we also mask channels that do not conform to the noise properties predicted by the radiometer equation in Eq. (3.6).

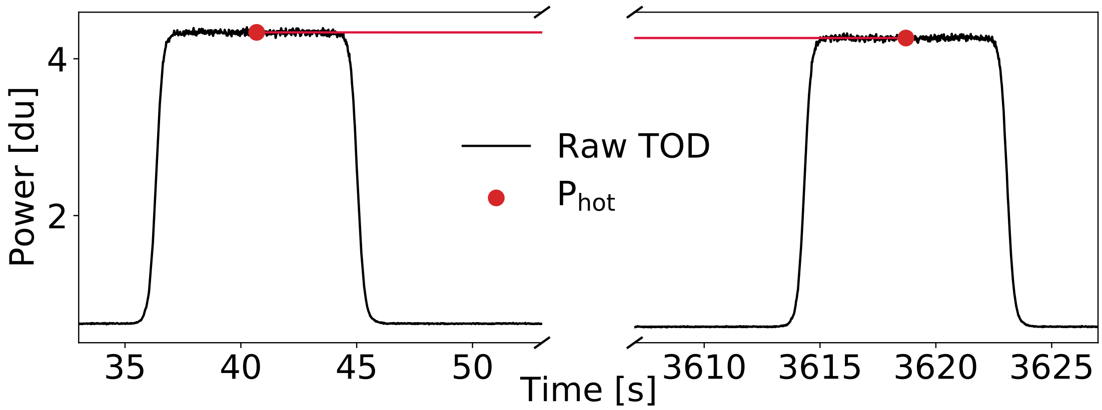


Figure 28: Reference load signal at the beginning and end of an observation introduced by rotating the calibration vane, seen in Fig. 13, in front of the detector array. Courtesy of Jonas Lunde, see Paper I.

After identifying bad channels, the mask is reapplied to the data before the poly- and time-domain PCA filters, and the filters are repeated (see loop-back in Fig. 23). In this way, we ensure that the identified channels do not affect the poly- and time-domain PCA filters and further analysis.

Calibration and system temperature measurement

After filtering and masking, the data are still in dimensionless normalized units. However, to perform physical inference with our data, they must be in calibrated brightness units.

Specifically, we want the TOD to be in units of the system temperature, T_{sys}, such that the changes in the data directly reflect the changes in the sky signal strength:

$$\Delta T_{\mathrm{sys}} = \Delta T_{\mathrm{signal}}. \tag{3.19}$$

The first step in this calibration is to find the system temperature at any given time. We do this by comparing the cold load power P_{cold}, measured at any given time during an observation of the sky, to a reference signal P_{hot} of known temperature T_{hot}. This reference hot load is measured twice for every observation by rotating the calibration vane (seen in Fig. 13) in front of the detector array. Figure 28 shows two such hot load measurements at the beginning and end of an observation. With the two calibration points at the beginning and end of any obsID, we can account (linearly) for changes in the ambient conditions around the telescope that can lead to drift in the system temperature.

From the cold and hot loads, P_{cold} and P_{cold}, we can obtain a T_{sys} measurement

$$T_{\mathrm{sys}} = \frac{T_{\mathrm{hot}} - T_{\mathrm{CMB}}}{\frac{P_{\mathrm{hot}}}{P_{\mathrm{cold}}} - 1}, \tag{3.20}$$

where T_{CMB} is the CMB temperature and T_{hot} is the ambient temperature measured at the calibration vane. As only two hot loads are taken per observation, we obtain a T_{sys} estimate at any scan in-between by linear interpolation.

The data are then calibrated by multiplying the filtered and masked data $d_{t\nu i}$, at time t, channel ν and detector i, by the corresponding average system temperature during the scan

$$d_{t\nu i}^{\text{Kelvin}} = d_{t\nu i}\langle T_{t\nu i}^{\text{sys}}\rangle. \tag{3.21}$$

For more details and derivations of the calibration, see Penzias & Burrus (1973), Lamb et al. (2022), and Paper I.

Downgrading

After filtering, masking, and calibration of the TOD, the data $d_{t\nu i}^{\text{Kelvin}}$ have the native 2 MHz frequency resolution of the spectrometer. However, it is not strictly necessary to have such high-frequency resolution to perform inference on the data at a later stage. Additionally, the large data volume required for native 2 MHz resolution demands a lot of disk space system and processing time. We can, therefore, reduce the number of high-resolution channels $n_\nu^{\text{high-res}}$ by coadding neighboring channels of the data d_ν^{high} so that the decimated data are given by

$$d_{\nu'}^{\text{low}} = \frac{\sum_{\nu\in\nu'} w_\nu d_\nu^{\text{high}}}{\sum_{\nu\in\nu'} w_\nu}, \tag{3.22}$$

where we typically sum over 16 high-resolution channels ν that are contained within a low-resolution channel ν'. The weights w_ν are given by the inverse variance $w_\nu = 1/\sigma^2_{\text{wn},\nu}$, where we compute the white noise level σ_{wn} at any given channel by

$$\sigma^2_{\text{wn}} = \frac{\text{Var}(d_t - d_{t-1})}{2}. \tag{3.23}$$

The resulting data $d_{\nu'}^{\text{low}}$ is now sampled on a 31.25 MHz grid instead of the previous 2 MHz spacing.

3.4.3 Mapmaking

Up until now, all analysis has been performed in the time domain. However, using the entire time-ordered dataset to perform physical inference would be unfeasible due to its enormous volume. We can, therefore, first bin the TOD data into maps, compressing the data from several hundreds of terabytes to only a few tens of gigabytes per map.

Generally, to get a map estimate $\hat{\boldsymbol{m}}$ from some time ordered data $\boldsymbol{d}$ with noise properties characterized by covariance N, and given the pointing matrix P we must solve the mapmaking equation (see Tegmark, 1997, for extensive explanation of mapmaking)

$$\mathsf{P}^T\mathsf{N}^{-1}\mathsf{P}\hat{\boldsymbol{m}} = \mathsf{P}^T\mathsf{N}^{-1}\boldsymbol{d}. \tag{3.24}$$

Assuming that the data are close to white noise after filtering, Eq. (3.24) can simplified by inserting a diagonal covariance matrix $\mathsf{N}_{ii} = \sigma^2_{\text{wn,i}}$. Equation (3.24) then reduces to a simple inverse noise-weighted binning

$$\hat{m}_p = \frac{\sum_{t\in p} d_t \sigma^{-2}_{\text{wn,t}}}{\sum_{t\in p} \sigma^{-2}_{\text{wn,t}}}, \tag{3.25}$$

where all the time samples t that hit a pixel p are accumulated. This is done individually for each frequency and feed and produces a 3D map with dimensions corresponding to right ascension, declination, and spectrometer frequency. Each element in the map, a so-called voxel or, in the corresponding cosmological units, a spaxel, is then made up of an (angular) pixel and a spectral channel. The corresponding uncertainty of the pixels in the map is then simply given by

$$\sigma_{\mathrm{wn},p} = \sqrt{\frac{1}{\sum_{t\in p} \sigma_{\mathrm{wn,t}}^{-2}}}. \tag{3.26}$$

Figure 29 shows an example of the three main COMAP Fields 1-3 as mapped out at the time of the COMAP ES results presented in Paper I. One can see that the central regions of the maps, which are observed more infrequently than the edges, are less noisy. This follows the prediction of the radiometer equation and Eqs. (3.25) and 3.26. Additionally, we can see from the maps, as well as the histograms over their uncertainty-normalized 3D voxels, that the ES maps appear to be dominated by white noise and follow a standard Gaussian $\mathcal{N}(0, 1)$ distribution.

3.4.4 New systematics discovered in COMAP S2 and the map-domain PCA filter

While the COMAP ES maps were found to be dominated by white noise, as we can see in Fig. 29, a series of new systematic effects manifested in the S2 maps due to a much higher sensitivity (more details of the increased data volume and data selection in Sec. 3.5). Specifically, we can divide the new systematics into two categories, which were dubbed the *turn-around* and the *start-of-scan* effects. Both of these effects are strongly correlated with both the pointing and frequency observed and are, as such, challenging to deal with. In the following, we will summarize these effects, show some examples thereof, and how we can filter them out as done in Paper III.

The turn-around effect

The turn-around effect manifests itself in the map as bright stripes at the declination edges of the map. Here, the telescope is at the edge of each CES sweep in azimuth, as illustrated in Fig. 16. In the frequency dimension, the excess seems to be slowly oscillating. In addition to the sharp excess at the edges of each scan, there are diffuse patterns that are roughly constant along fixed declinations in the middle of the fields. However, note that the diffuse contribution could also simply be explained by ringing when filtering the data with sharp peaks at the turn-around edges. The leading hypothesis for the origin of the turn-around effect is mechanical vibrations during the scanning motion of the telescope. This seems to be supported by the fact that the effect was apparently somewhat dampened after the scanning speed of the telescope was reduced about mid-way through S2 (named Season 2b in Paper III).

In the first and third row of Fig. 30, we show some examples of the turn-around effect in the S2 Field 2 map of different feeds and frequencies. Note that the systematic effects seen in Fig. 30 appear twice for every map because the field is observed both while it is rising and setting on the sky, giving slightly different line-of-sights each time.

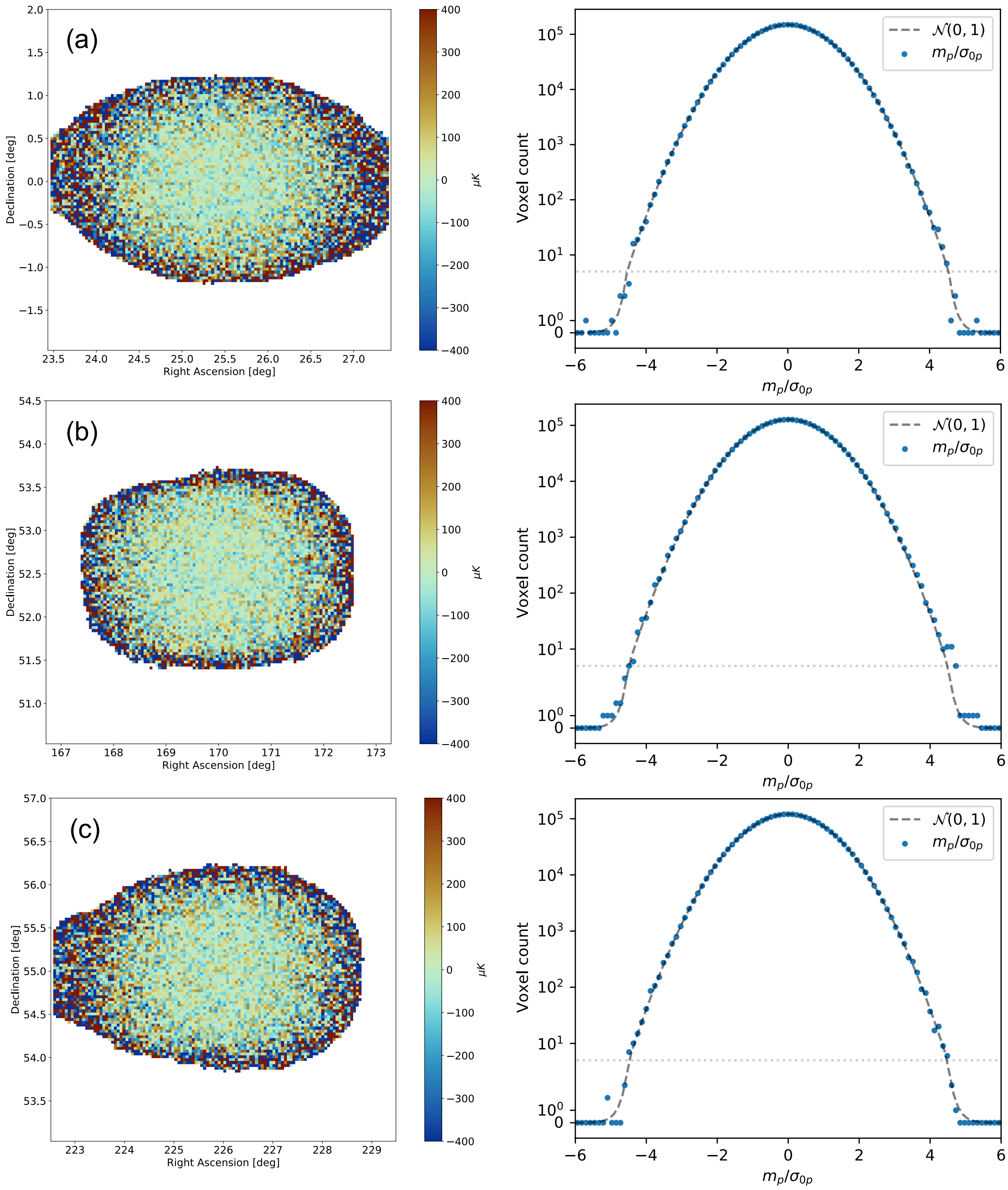


Figure 29: Single 31.25 MHz frequency channel and feed-coadded map from COMAP Early Science (Paper I) at 29.9 GHz of the three fields 1-3 (a-c). Note that regions with high noise, $\sigma_{\rm wn} > 1000\,\mu$K, are masked from the plot. A histogram of the uncertainty-normalized voxel values of each entire 3D map is shown next to the map frequency slice along with a standard normal $\mathcal{N}(0,1)$ distribution.

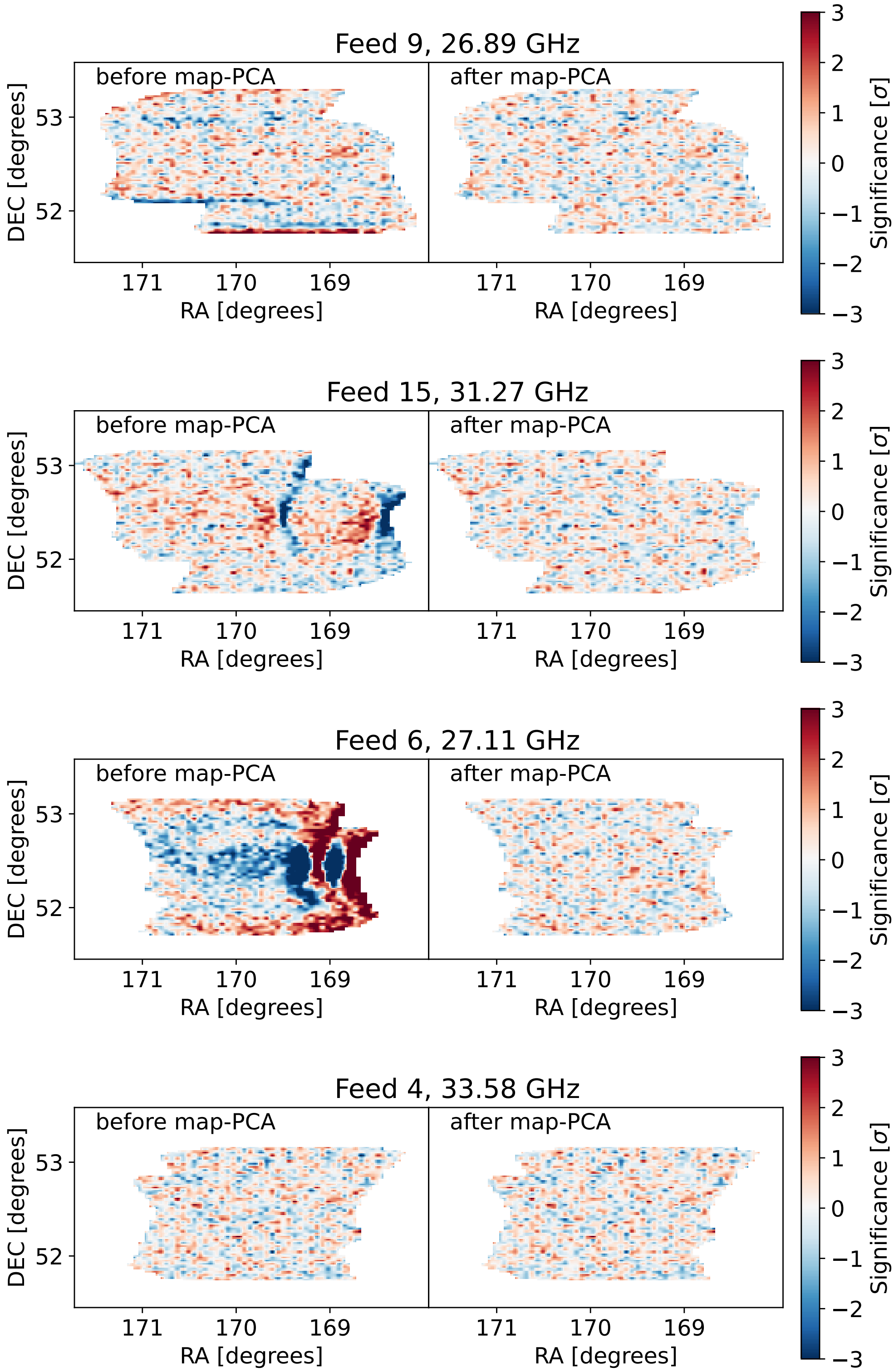


Figure 30: Season 2 maps of Field 2 at different feeds and frequencies in units of their voxel uncertainty. The maps show, respectively from top to bottom, the turn-around effect, the start of scan effect, both of the latter at the same time, or neither effect. The left and right columns show the maps before and after the map-PCA filter subtracts the leading five modes. Courtesy of Jonas Lunde, see Paper III.

The start-of-scan effect

The start-of-scan effect, as the name suggests, is a systematic effect that appears at the beginning of scans. Specifically, it appears to behave like a standing wave in frequency but has an exponential decay in time of 19 s on average. Consequently, the effect always appears as a bright excess (with the sign depending on the frequency considered) at the lower RA of the maps. This coincides with where the telescope points at the beginning of each scan, both when the field rises and sets in the sky. The start-of-scan effect is illustrated in the second and third rows of Fig. 30, and we can see from Fig. 16 how a scan always starts at lower RA. Note that when going towards increasing RA the sign of the start of scan amplitude seems to flip. This is due to the ringing of the normalization step described in Sec. 3.4.2.

The fact that the effect manifests only at the start of a scan suggests that the cause for the observed decaying standing wave is the settling time of mechanical vibrations in the telescope after repointing itself between scans. Additionally, the start-of-scan effect is primarily observed in feeds 6, 14, 15, 16, and 17, which all share the same DCM1 (DCM1-2; see Table 3), meaning that a likely culprit causing this common mode signal would be common local oscillator cables shared within DCM1-2.

The map-PCA filter

Because the turn-around and start-of-scan effects appear to be fairly weak in strength, they appear only after combining several significant months of data. It is, therefore, difficult to model and remove in the time domain, where the S/N of the effect is very low. However, the effects are orders of magnitude stronger than the extragalactic CO in the maps and are well modeled by a standing wave. In such a case, where we have a hard-to-model and dominant signal that is decomposable into an outer product of a few standing wave modes, a PCA approach is well suited. Meanwhile, the extragalactic signal is weak and not easily separable into an outer product, which is an additional safeguard against signal loss. Methods based on PCA or other blind PCA-like techniques have also been used previously by the H I LIM community to remove Galactic continuum foreground emission from their maps (Chang et al., 2010; Masui et al., 2013; Wolz et al., 2017; Anderson et al., 2018; Cunnington et al., 2021).

Specifically, we can write down a similar PCA minimization problem as previously shown in Eq. (3.17). Only this time it is applied to the map $m_{p\nu}$, with n_{pixel}^2 pixel p, instead of the TOD $d_{t\nu}$ at time t. Additionally, we want to use weights so the PCA does not simply pick up the inherently noisy regions of the map. We thus must minimize

$$\sum_{p,\nu} \frac{(m_{p\nu} - w_p^0 a_\nu^0)^2}{\sigma_{p\nu}^2}, \tag{3.27}$$

where $\sigma_{p\nu}$ is the voxel dependent noise level. To find the vectors w_p^0 and a_ν^0 that explain most of the variance inside the map data matrix $m_{p\nu}$. Note that the minimization in Eq. (3.27) differs somewhat from the previous PCA shown in

Eq. (3.17) because the weights applied are not separable as an outer product themselves. Thus, the problem of minimizing Eq. (3.27) technically becomes a modified PCA that is not directly equivalent to applying a PCA to the weighted map $m_{p\nu}/\sigma_{p\nu}$ (this is explained in detail in Appendix B of Paper III).

After finding the map-PCA components w_p^0 and a_ν^0 we can filter the map by

$$m_{p\nu}^{\mathrm{mPCA}} = m_{p\nu} - w_p^0 a_\nu^0. \tag{3.28}$$

To remove further map-PCA components, we simply iterate over the minimization and subtractions shown in Eqs. 3.27 and (3.28). Typically, most feeds seem to require only 3–5 of the 256 possible PCA modes subtracted to suppress systematic errors below the noise level. In the S2 results (Papers III–V), we subtract the five leading components of each map. Figure 30 shows the maps before and after subtracting the five leading map-PCA modes. We see that the strong systematic effects observed before filtering are now no longer visible above the noise level. See Paper III for more results on the map-PCA's effect on the data.

Because we see quite different systematic effects in different feeds, the filter is applied individually to each detector map. We also perform the map-PCA filter independently on the maps of the fast- and slow-moving azimuth scans because the mechanical vibrations, and hence the systematic effects they cause, can be different.

Lastly, we note that because a PCA is a nonlinear operation it can affect the signal in the map in unpredictable ways that are hard to quantify. However, as long as the CO signals of interest are well below some critical S/N, the map-PCA behaves approximately linearly with respect to the signal. We show this using simulations with varying S/N and compute the signal loss transfer function (more on this later in Sec. 3.6.3) after applying the map-PCA. The result, seen in Fig. 31, indicates that the critical S/N in our case is at about 0.02 for an average map-voxel. This is about an order of magnitude higher S/N than currently achieved by COMAP S2 (blue line), putting us firmly in the linear signal regime of the map-PCA.

3.5 Data selection

An additional step used to prevent systematic effects from affecting final data products is data selection. That is, if data with significant residual systematic effects are identified and removed, they cannot negatively affect further data analysis. Therefore, data selection is most effective when performed in the earliest possible stage. However, in the earlier pipeline stages, identifying bad data can be harder since the S/N ratio is low when individual scans are considered separately. To get the best of both worlds, we can perform data selection in several stages of increasing S/N: when observing, in time- and map-domain analysis, and when computing power spectra. For each data selection step, we can then estimate the data retention (that is, how much of the full data are kept in the analysis). The data retention from COMAP ES and S2 can be found in Table 2 (from Paper III) and has improved significantly in the latest publications. In the following, we will briefly summarize the data selection steps.

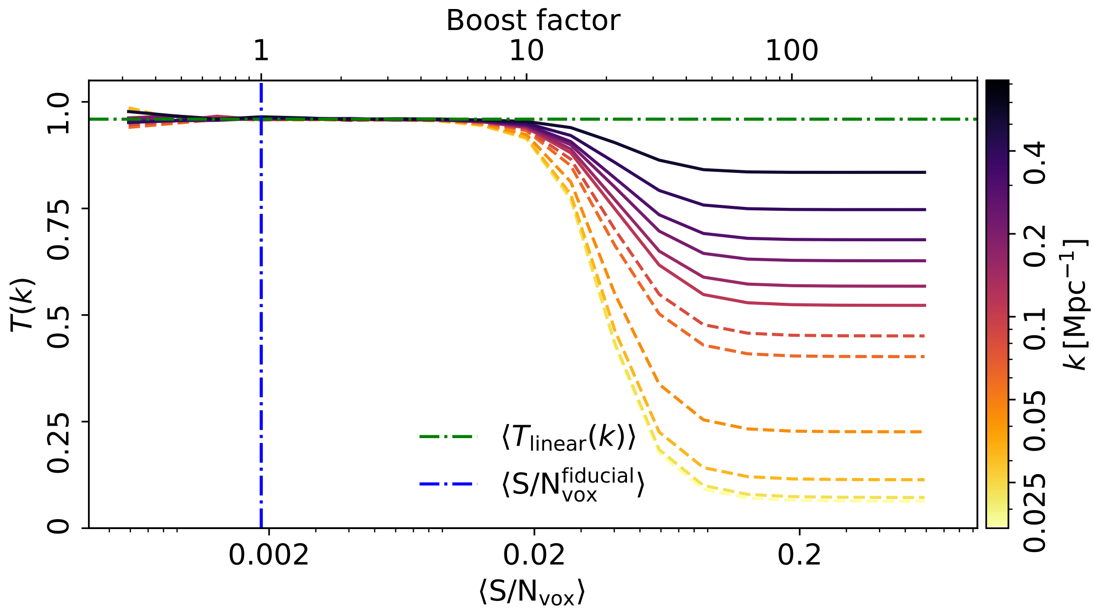


Figure 31: Signal transfer function $T(k)$ of the map-PCA filter resulting from filtering simulated signal maps as a function of varying average voxel signal-to-noise ratio (x-axis) and at different scales k (color scale). For each S/N the signal strength with respect to the fiducial CO model by Chung et al. (2022a) (marked as a blue line at unity boost) is shown in the upper x-axis. The average transfer function, $\bar{T}(k) \approx 0.96$, at the noise-dominated low-S/N range is marked as a green line. Scales, k, beyond the range of interest for the S2 publications (Paper III-V) show dashed lines for $T(k)$. The figure is taken from Paper III.

Observational cuts

The earliest stage of data selection is performed when the sky is observed. As mentioned earlier, the scans of Season 1 that used a Lissajous scanning strategy were found to contain residual excesses and were therefore rejected from the analysis. Learning from this lesson, S2 only employed the CES scanning strategy. Furthermore, there were several feeds that did not function as intended or that were used for engineering tests in Season 1, whereas all feeds were functional and included in S2. Lastly, as mentioned earlier, the elevation range at which to observe was slightly reduced in S2 to avoid areas with sharp ground pickup signal gradients. Therefore, as seen in the upper third of Table 2, the data retention from these three cuts was $E_{\text{obs}} = 0.318 = 0.50 \cdot 0.842 \cdot 0.755$ in Season 1 and increased to 100 % in S2.

Time- and map-domain cuts

Several data selection cuts are also performed in the time- and map-domain. The first of these, frequency masking in `l2gen`, was already described in detail in Sec. 3.4. Data retention from frequency masking E_{freq} has remained fairly constant between Season 1 and S2, even though the frequency masks were changed. Although the data of S2 with an increased spectrometer sampling frequency have

	Season 1	Season 2	Explanation
E_{scan}	50.0%	100.0%	Retained scans (CESs)
E_{feed}	84.2%	100.0%	Functional feeds
E_{el}	75.6%	100.0%	Inside good elevation range
$\boldsymbol{E_{\text{obs}}}$	**31.8%**	**100.0%**	Observational data retention
E_{freq}	72.8%	74.3%	Frequency masking in `l2gen`.
E_{stats}	57.4%	33.6%	Cuts on accept-mod statistics
$E_{\chi^2_{P(k)}}$	72.2%	100.0%	Per-scan auto-PS χ^2 test
$\boldsymbol{E_{\text{cuts}}}$	**30.1%**	**24.9%**	Map-level data retention
$E_{\chi^2_{C(k)}}$	52.4%	100.0%	Cross-spectrum χ^2 test
$E_{C(k)}$	94.7%	75.0%	Cross-spectrum auto combinations
$\boldsymbol{E_{\text{PS}}}$	**49.6%**	**75.0%**	Retained data at PS-level
$\boldsymbol{S_{\text{tot}}}$	**6.8%**	**21.6%**	Final PS-domain sensitivity, calculated as $S_{\text{tot}} = \sqrt{E^2_{\text{obs}} E^2_{\text{cuts}} E_{\text{PS}}}$

Table 2: Overview of data retained at different stages of the analysis pipeline, for both S1 (ES; Papers I and II) and S2 (Papers III-V). The table is divided into three. From top to bottom, we find the data retention for the observational, time- and map-domain analysis, and power spectrum data selection. Each individual step has its associated data retention fraction E_i, and the last row of each sub-table contains the combined product of the data efficiencies, E_{obs}, E_{cuts} and E_{PS}. The very last row shows the total data efficiency, S_{tot}, as a fraction of the theoretical maximum. The table is taken from Paper III, courtesy of Jonas Lunde.

no aliasing cut at the edge of each band, several more frequency channels that showed evidence of systematic effects were manually removed. This balances out the data volume gained from loosening the aliasing cuts.

After filtering and calibrating the data in `l2gen`, but before mapmaking, several more time-domain data selection steps are performed in the `accept_mod` code. First, `accept_mod` takes in all scans of a field and makes a database of diagnostics for each scan, feed, and sideband. The database contains both statistics (about 60–70 depending on the COMAP release) related to the scanning strategy (e.g., mean elevation and azimuth, whether the moon or sun is in one of the sidelobes), the telescope environment (e.g., weather, humidity or wind speed), and pipeline variables (e.g., the number of PCA modes subtracted, the f_{knee} of the polyfilter coefficients c_i of Eq. (3.13) in each time sample). Using the scan database, we can then reject scans, feeds, or sidebands with diagnostics outside of some predefined range. The data retention from this stage is referred to as E_{stats} in Table 2. Because the scan diagnostics data selection in S2 was significantly stricter, in order to pass later power spectrum null tests (see Sec. 3.6.4, the data retention E_{stats} dropped between the two seasons.

Among the diagnostics computed by `accept_mod`, there are several χ^2 statistics computed from 2D and 3D auto power spectra of low-resolution maps of the data. These are meant to detect excesses from the expected white-noise behavior in each scan. However, the method requires us to adjust the power spectrum by a transfer function to account for the filter-induced bias on the noise, which proved hard to quantify for individual scans, fields, and scanning speeds due to very different hit patterns. Additionally, while working on the S2 results in Papers III and V, it seemed as if these cuts had little effect on the data, and therefore, we deemed it unlikely that they identified any significant and dangerous systematic effects. As a result, the data retention $E_{\chi^2_{\mathrm{P(k)}}}$ from these cuts are at 100 % in S2, while only 72.2 % in Early Science.

In total, scan-based data selection leaves us with data retention $E_{\mathrm{cuts}} = 30.1\,\%$ in the Season 1 analysis, while the stricter scan diagnostics cuts in S2 dropped E_{cuts} to 24.9 %.

Power spectrum data cuts

The final data selection stage is performed at the power spectrum level. As will be described later in Sec. 3.6, the COMAP power spectrum methodology is based on computing cross-power spectra between feeds (in Paper II), or feed groups (Paper IV), and scan elevations. Before combining all feed-elevation cross-spectrum combinations into an average CO power spectrum estimator, a χ^2 test can be performed on each combination. Subsequently, we reject cross-spectra that are more than 5σ away from the expectation value of zero. However, this cut was removed in S2 in favor of a more robust null test framework and more strict data cuts at the scan level. This essentially doubled the retention of the data $E_{\chi^2_C(k)}$ seen in Table 2.

In Seasons 1 and 2 we computed, respectively, average cross-power spectra from feeds and groups of feeds, where auto-detector combinations are excluded from the average. Therefore, we lost 19 of the 361 possible cross-spectra in Season 1. In contrast, the S2 feed-group cross-spectra result in 16 total combinations, of which four auto-combinations are rejected. This gives an inherent data retention $E_{C(k)} = 94.7\,\%$ and 75 % of the two methods (and seasons). Combining the two power spectrum cuts the power spectrum level data retention E_{PS} is 49.6% and 75 % in the two COMAP seasons, respectively.

Fraction of maximum possible sensitivity

If we want to combine the data retentions E_i of Table 2 into a final number to determine how much more sensitive the power spectrum of S2 is compared to that of S1, we cannot simply multiply all individual data retentions. This is because the power spectrum sensitivity scales linearly with the raw data volume while it scales as the square root of the number of combined power spectra. We thus get a total fraction of the maximally achievable sensitivity $S_{\mathrm{tot}} = \sqrt{E^2_{\mathrm{obs}} E^2_{\mathrm{cuts}} E_{\mathrm{PS}}}$.

As seen in Table 2, the number of loosened cuts and the optimization of the COMAP observational strategy has dramatically increased the total sensitivity fraction S_{tot} from 6.8 % to 21.6 %, despite some cuts being tightened. In other words, there is a factor of roughly 2.2 increase in maximum power spectrum sensitivity retention. Combining this with the 3.4 times increased raw data volume in the combined analysis of the Season 1 and 2 data, the power spectrum sensitivity based on the data volume and data selection is expected to be around 7 times higher than only the Season 1 power spectrum we present in Paper II.

3.6 The high-level analysis pipeline

A key goal of any cosmological and astrophysical survey is to constrain the physics of the mapped environment. To perform inference, we must find out how the signal in the maps reacts to changes in the underlying physics. This can be done by first constructing a summary statistic that is sensitive to the statistical information stored in the map and then performing inference on the summary statistic of the map.

A popular summary statistic, often used by CMB and large-scale structure surveys, is the power spectrum[3]. Power spectra encode all statistical information in the signal if it is Gaussian randomly distributed and isotropic. However, LIM surveys probe a wide range of physical scales and environments, from large linear scales to small nonlinear scales where Gaussianity no longer holds. In cases where the probed field is no longer Gaussian, the power spectrum can be complemented with higher-order statistics such as the bi- and trispectrum (Yoshiura et al., 2015; Shimabukuro et al., 2016; Planck Collaboration IX., 2020; Planck Collaboration X., 2020), or one-point statistics like the voxel intensity distribution (VID; Breysse et al., 2017; Ihle et al., 2019) and the deconvolved distribution estimator (DDE; Breysse et al., 2023; Chung et al., 2023). Another high-level analysis technique that can be used in inference is to obtain the average LIM signal from stacking on galaxy positions from an external galaxy catalog (Dunne et al., 2024, 2025).

Nevertheless, there are not yet any robust auto-correlation LIM detections. Therefore, supplementing the power spectrum with other summary statistics becomes somewhat moot. In this section, we summarize and tie together the work presented in Papers II, IV, and V, respectively covering the power spectrum methodology and results of COMAP ES, COMAP S2, as well as the physical inference of COMAP S2. Helping to develop and further improve the COMAP power spectrum methodology constitutes the largest contribution to the combined work presented in this thesis.

3.6.1 Power spectrum analysis

We can divide power spectra into two types. The first of these, the *auto-power spectrum*, is equivalent to the variance of the Fourier coefficients in a map. We can write the auto-power spectrum as

$$P(\boldsymbol{k}) = \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}}}\langle|\mathcal{F}\{\boldsymbol{m}\}(\boldsymbol{k})|^2\rangle = \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}}}\langle|f(\boldsymbol{k})|^2\rangle, \tag{3.29}$$

where the voxel volume and number of voxels are given by V_{vox} and N_{vox}, and the Fourier coefficients $f(\boldsymbol{k})$ are given by the Fourier transform of the map $\mathcal{F}\{\boldsymbol{m}\}(\boldsymbol{k})$. Because the COMAP maps $\boldsymbol{m}$ are a function of a three-dimensional position in

[3] The Fourier equivalent of the real-space two-point correlation function.

DCM1 (feed group)	Feeds
1	1, 4, 5, 12, 13
2	6, 14, 15, 16, 17,
3	2, 7, 18, 19
4	3, 8, 9, 10, 11

Table 3: "Feed groups" and their associated first down-converter module (DCM1). See Fig. 13 for a visual representation of which feeds belong to a certain DCM1.

(real-)space (often measured in comoving Mpc), the power spectrum must be a three-dimensional function of the wavenumber vector $\boldsymbol{k}$ (in units Mpc^{-1}).

By virtue of being the variance of the map's Fourier coefficients, all sources of variance in the map will contribute to the auto-spectrum

$$\langle P(\boldsymbol{k})\rangle = P_{\mathrm{CO}}(\boldsymbol{k}) + P_{\mathrm{syst}} + P_{\mathrm{noise}}(\boldsymbol{k}), \tag{3.30}$$

as the sum of the signal, systematic effects, and noise power spectra.

Similarly, we can also define the so-called *cross-power spectrum* as the covariance between Fourier coefficients $f_i(\boldsymbol{k})$ and $f_j(\boldsymbol{k})$ of two maps $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$. This is then written as

$$C(\boldsymbol{k}) = \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}}}\langle \mathrm{Re}\{f_i^*(\boldsymbol{k})f_j(\boldsymbol{k})\}\rangle, \tag{3.31}$$

where "$*$" denotes complex conjugation. We see that this reduces to the auto-power spectrum $P(\boldsymbol{k})$ in Eq. (3.30) if the two maps $\boldsymbol{m}_i = \boldsymbol{m}_j$.

In contrast to the auto-power spectrum $P(\boldsymbol{k})$, the cross-spectrum $C(\boldsymbol{k})$ will only be sensitive to common contributions to the maps $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$. This means that if the two maps have both independent noise and systematic effects but share the same CO signal, the cross-spectrum of the maps can be written as

$$\langle C(\boldsymbol{k})\rangle = P_{\mathrm{CO}}(\boldsymbol{k}). \tag{3.32}$$

Meanwhile, to obtain the same signal estimate through an auto-power spectrum like in Eq. (3.30) we need to model both the systematic effects and noise to a high precision to not bias the signal estimate $P_{\mathrm{CO}}(\boldsymbol{k})$. It is still important to know the systematic effects and noise properties of the maps when estimating a cross-spectrum as in Eq. (3.31), but only to obtain accurate uncertainty estimates and not to obtain the correct expectation value.

In practice, because our maps have highly nonuniform noise properties, we need to weigh the maps prior to computing Fourier transforms to prevent the power spectrum from being dominated by high-noise regions of the map. The resulting *pseudo-power spectra* will be a biased estimator of the signal power spectrum as different Fourier modes are coupled through the applied weights (see more on this mode-mixing in Hivon et al., 2002; Leung et al., 2022, and Appendix D of Paper II).

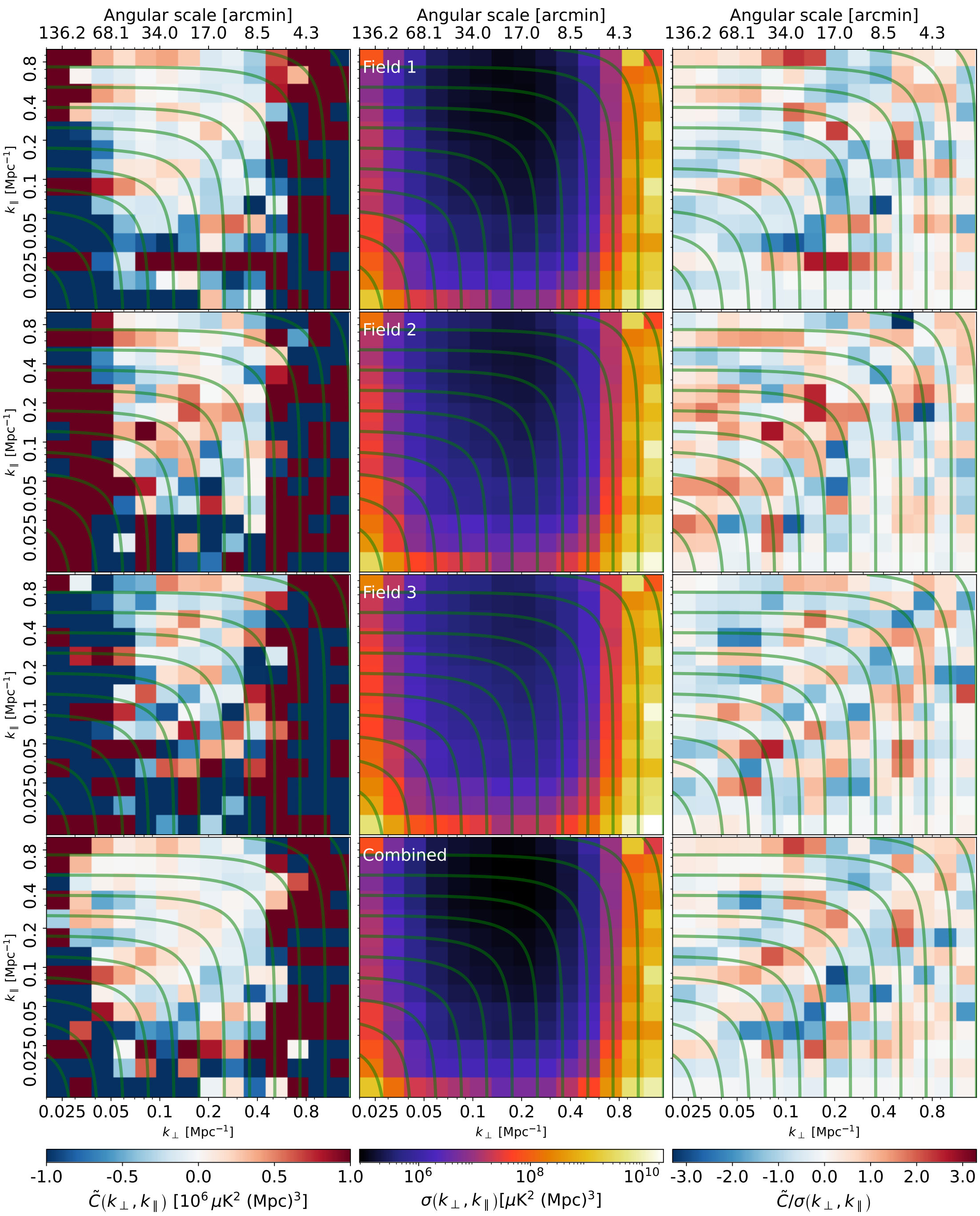


Figure 32: (*Left column*) Cylindrically averaged power spectra of the full COMAP Early Science data (Paper II) for each of the three COMAP fields (*first three rows)* as well as coadded over fields (*last row*). (*Middle column*) Corresponding uncertainty estimates and (*right column*) the ratio between power spectra and their uncertainties. Green contours indicate k-bin edges of the spherically averaged power spectrum seen in Fig. 43.

3.6.2 From feed-feed to feed group pseudo-cross-power spectra

The sensitivity and robustness against systematic effects and noise assumptions make the cross-spectrum an ideal starting point for the COMAP power spectrum methodology, as described in Papers II and IV. Specifically, the methodology builds on cross-correlating different feeds or feed groups and elevation maps to eliminate the noise bias and systematic contamination from the final CO power spectrum estimate (as shown in Eq. (3.31)). This is done in the `comap2fpxs` step of the COMAP pipeline in Fig. 23.

Because each feed has a (largely) independent detector chain and associated electronics, they also have independent noise and (largely independent) systematic effects. A worrisome systematic effect for COMAP is pointing-correlated far-sidelobe ground pickup from the mountainous horizon at Owens Valley, which may not average down with more data due to the repeated motion of the telescope across the sky. To avoid ground pickup, we also want to cross-correlate between maps of only high- and low-elevation scans, each with different ground pickup, as we can see from Fig. 22.

This leads us to the feed-feed pseudo-cross-power spectrum (FPXS; described in detail in Paper II), as well as the feed group pseudo-cross-power spectrum (FGPXS; see Paper IV). Both cross-correlate data across detectors and elevations to cancel the noise bias, as well as detector- and elevation-specific systematic errors. The only difference between the two methods is that the FPXS cross-correlates between individual feeds, while the FGPXS relies on cross-spectra between certain feed configurations that share systematic effects of the type seen in Sec. 3.4.4. Otherwise, the FPXS and FGPXS remain completely analogous. Specifically, feeds with a common DMC1, as shown in Table 3 and Fig. 13, are grouped together to avoid cross-correlating feeds with shared systematic effects. Additionally, the larger cross-map footprint of feed group maps can help prevent larger mode mixing and, therefore, leakage of large-scale systematics effects into the small and intermediate scales of interest.

We can, therefore write both methods algorithmically as follows:

1. Divide the data into two halves A and B. As mentioned in Papers II and IV, we chose to split the data into high- and low-elevation scans to isolate different ground signals.

2. For each individually processed half A and B make feed (group) maps and let them be denoted as, e.g., $\boldsymbol{m}_{A_2}$ for maps of elevation A and feed (group) 2.

3. Compute all possible (pseudo-)cross-spectrum combinations $C_{A_iB_j}(\boldsymbol{K})$ between maps $\boldsymbol{m}_{A_i}$ and $\boldsymbol{m}_{B_j}$.

4. Combine all $N_{\text{feedgroup}}(N_{\text{feedgroup}}-1)$ computed cross-spectra, $C_{A_iB_j}(\boldsymbol{K})$, for which $A \neq B$ and $i \neq j$ so that

$$C(\boldsymbol{k}) = \frac{\sum_{i\neq j, A\neq B} \frac{C_{A_iB_j}(\boldsymbol{k})}{\sigma^2_{A_iB_j}(\boldsymbol{k})}}{\sum_{i\neq j, A\neq B} \frac{1}{\sigma^2_{A_iB_j}(\boldsymbol{k})}}, \tag{3.33}$$

to obtain the mean feed-feed or feed group pseudo-power spectrum (FPXS and FGPXS).[4] Depending on whether FPXS or FGPXS are computed, $N_{\text{feedgroup}} = 19$ or $N_{\text{feedgroup}} = 4$.

Using Eq. (3.33) we can compute combined CO power spectrum estimates for each separately processed field. In Figs. 32 and 33, we see an example of the main cylindrically averaged FPXS and FGPXS results adapted from Papers II and IV. These contain the CES data of Season 1 and (cumulatively) S2. As one can see, the noise on small and large scales is blown up. This is due to the pipeline filters, voxel windows, and beam transfer functions, which we will return to in Sec. 3.6.3. The main goal of the ES results was to show that the COMAP analysis pipeline removes systematic effects down to the white noise level so that the noise can average down with more future data. As we see from Fig. 32, this seems to be the case up to some correlated fluctuations (more on these in Sec. 3.6.2) at the smallest $k_\perp$. These could be explained by standing wave residuals or poorly constrained modes from suboptimal cross-feed sky overlaps. We can also see in the right columns of Fig. 32 that the noise in the data is somewhat underestimated at low $\boldsymbol{k}$ and vice versa at intermediate and high $\boldsymbol{k}$. This is due to the way the power spectrum uncertainties are estimated in Paper II, and we will come back to this in detail in Sec. 3.6.4 as well as how it is improved in Paper IV.

Meanwhile, in Fig. 33, we show the entire $\boldsymbol{k}$-space of the COMAP S2 power spectra from Paper IV, including the masked regions. We can see that the power spectra on small and intermediate scales seem to be noise-dominated. We also see that the noise level does not seem over- or under-estimated. However, we see clear signs of systematic effects on large perpendicular scales, which motivates the COMAP S2 $\boldsymbol{k}$-mask, and we come back to masking later in Sec. 3.6.2.

Sensitivity of an FGPXS compared to an FPXS

As we see from Eq. (3.33), of all possible combinations between feeds and elevation, only those that neither cross the same feed nor elevation are propagated to the final average cross-spectrum. This is graphically illustrated for both FPXS and FGPXS in Fig. 34. When crossing individual feeds instead of groups of feeds a larger percentage of the area of the triangle in Fig. 34 is maintained and consequently the average feed cross-power spectrum will be more sensitive than the feed group equivalent. This can, as shown in Paper IV, also be expressed in terms of the fraction of the maximally achievable auto-power spectrum sensitivity $\sigma_{P(k)}$ (i.e, combining all cells in Fig. 34):

$$\frac{\sigma^{N_{\text{split}} N_{\text{feed}}}_{C(\boldsymbol{k})}}{\sigma_{P(\boldsymbol{k})}} \geq \sqrt{\frac{1}{1 - \frac{1}{N_{\text{split}}}}} \sqrt{\frac{1}{1 - \frac{1}{N_{\text{feed}}}}}, \tag{3.34}$$

where the first square-root always reduces to $\sqrt{2}$ because $N_{\text{split}} = 2$ in Papers II and IV. We can find a theoretical loss in sensitivity between the FPXS and FGPXS of around 12 %, ignoring cross-overlap between maps, etc.

[4] Note that the FPXS and FGPXS are in practice binned into 1D or 2D representations, i.e., as functions of $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ and $(k_\perp, k_\parallel) = (\sqrt{k_x^2 + k_y^2}, |k_z|)$, by spherical and cylindrical

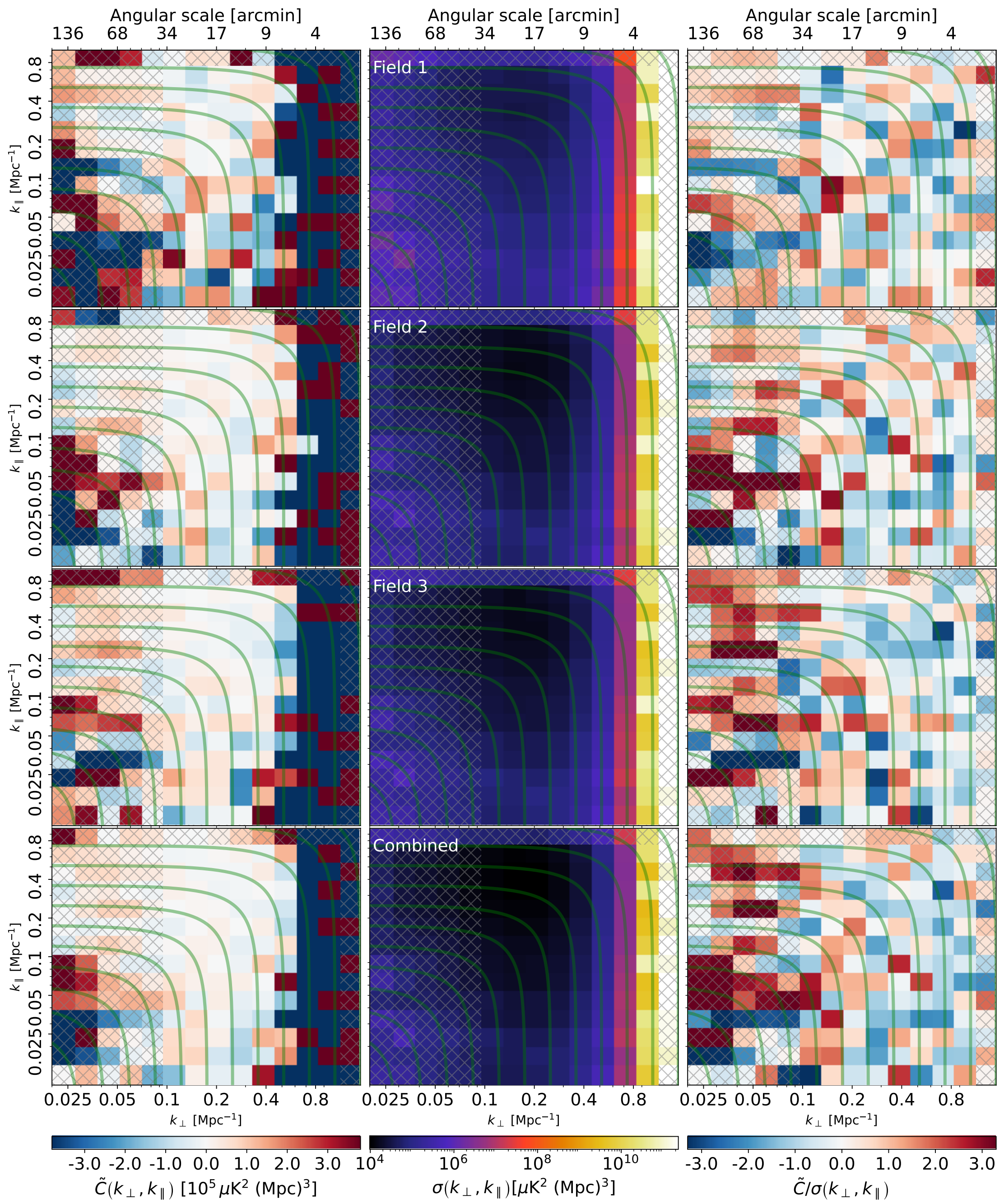


Figure 33: Same as Fig. 32 but using the full COMAP S2 data and $\boldsymbol{k}$-space mask. The figure is similar to Fig. 3 of Paper IV, but also shows masked areas (hatched).

However, in practice, we observe a somewhat higher loss in sensitivity, although the exact reasons are not yet properly understood and will be subject to future investigation. In Fig. 35, we see an example of the (unity subtracted) ratio between the uncertainty estimates, $\sigma_{\text{feedgroup}}$ and σ_{feed}, for an FGPXS and FPXS (both estimated as described later in Sec. 3.6.4) of Field 3. Both the FPXS and FGPXS

averaging to form the main summary statistic used later in inference. See Papers II and IV for details on how and why the spectra are binned.

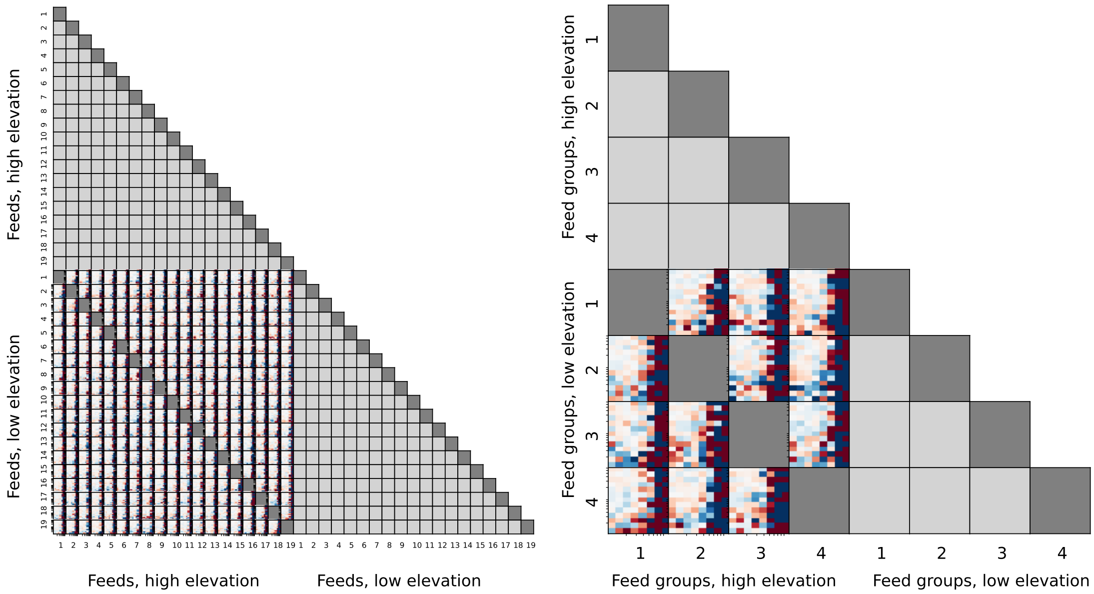


Figure 34: Grid of possible feed (*left*) and feed group (*right*) combinations for one of the COMAP field maps. Only combinations with 2D cross-spectrum data are used for a final average FPXS or FGPXS. In contrast, all the dark and light gray (auto-feed and auto-elevation combinations respectively) are excluded. The figure is adapted from Paper IV.

include the same raw data volume, and the only algorithmic difference is the cross-correlated feed configurations. On the very largest scales, which are masked in the COMAP S2 analysis, the FGPXS sensitivity is higher than that of the FPXS. This is most likely caused by the map weights, $w \propto 1/\sigma_1\sigma_2$, which are used in computing the pseudo-power spectra in Papers II and IV to make the mode mixing problem easily invertible at a later stage. The result, however, is that non-overlapping regions of the sky do not contribute to the cross-spectrum at all. Hence feed group maps, having slightly larger footprints, have a superior overlap on the very largest Fourier scales. Meanwhile, on smaller scales, the FPXS have a median of 18% lower uncertainties than the FGPXS, which is a somewhat larger difference than expected by Eq. (35). This will be subject to future investigations.

Power spectrum k-space mask

In COMAP ES (Paper II) all computed $(k_\perp, k_\parallel)$-bins are used for the final data products. However, in COMAP S2 (Paper IV, we had to include a $\boldsymbol{k}$-space mask to prevent any potential residuals of the turn-around or start-of-scan effects (Sec. 3.27) from affecting the analysis. In particular, data on perpendicular scales $k_\perp > 0.1\mathrm{Mpc}^{-1}$ are masked from the S2 analysis. This can be seen in, e.g., Fig. 33.

Furthermore, the systematic contamination on these scales can be tested further by performing null tests (see 3.6.4). These fail if the $k_\perp > 0.1\mathrm{Mpc}^{-1}$ mask is not applied, indicating that there are indeed some residual systematics on the very largest map scales. Additionally, we mask the outermost $k_\perp$- and $k_\parallel$-bins since these are closest to their respective Nyquist limits where aliasing problems

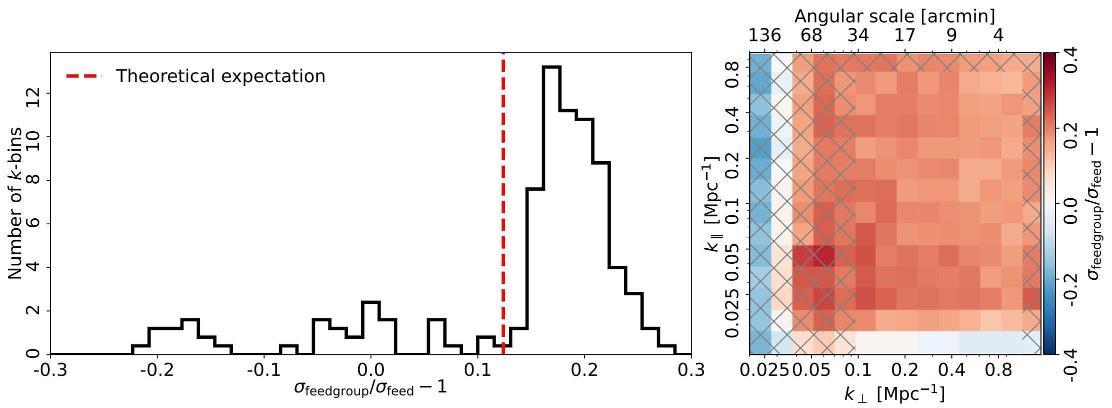


Figure 35: *Right*: Unity subtracted ratio between feed group and feed cross-spectrum sensitivity, $\frac{\sigma_{\mathrm{feedgroup}}}{\sigma_{\mathrm{feed}}} - 1$, COMAP Field 3. Gray hatching indicates the COMAP S2 mask. *Left*: Histogram of the right panel. The red line indicates the theoretical sensitivity loss of using feed groups instead of individual feeds as expected from Eq. (3.34).

are the strongest.

In Fig. 36, we show correlation coefficients between a selection of $\boldsymbol{k}$-bins and all the others. We note that on large scales, there are fairly strong correlations that can be as large as $\geq 30\,\%$, extending along constant $k_\parallel$. These correlations are a sign of poorly constrained large-scale modes where the maps lack effective sky overlap. We see that the correlations decrease towards small scales where the maps are dominated by noise. However, we see that the correlations become significantly less spread out in $k_\parallel$ outside of the discarded regime $k_\perp > 0.1\,\mathrm{Mpc}^{-1}$. Currently, these correlations are not properly taken into account when averaging feeds or binning the power spectra into spherical or cylindrical averages. Thus, properly addressing these correlations, especially when trying to recover the largest masked scales, will be an important future goal.

Finally, we note that by masking the large-scale regions of the maps, we also lose a significant portion of the area where COMAP is estimated to have a relatively high S/N. This can be seen in Fig. 37, where the masked region $k_\perp > 0.1\mathrm{Mpc}^{-1}$ is indicated by red hatching. The estimated normalized S/N assumes a Chung et al. (2022a) fiducial model. While the S/N is relatively high in the excised regions, a substantial part of the total S/N is kept in the COMAP S2 analysis.

Therefore, we err on the side of caution in Paper IV and mask the systematics contaminated, mode-mixed, and highly correlated regime at the price of losing some signal. That being said, as most of the clustering power (see Sec. 3.6.5) and significant amounts of the total S/N lay in the masked region of the power spectrum, recovering these scales will be an important goal for the future.

3.6.3 Power spectrum transfer functions

In a perfect experiment, a telescope has infinite resolution, the voxels of the maps are infinitesimally small, and no filtering has to be performed to clean the data. In such a case, the signal power spectrum estimator would simply be given by Eq. (3.32). However, in reality, this is never the case. That is, we have to take into

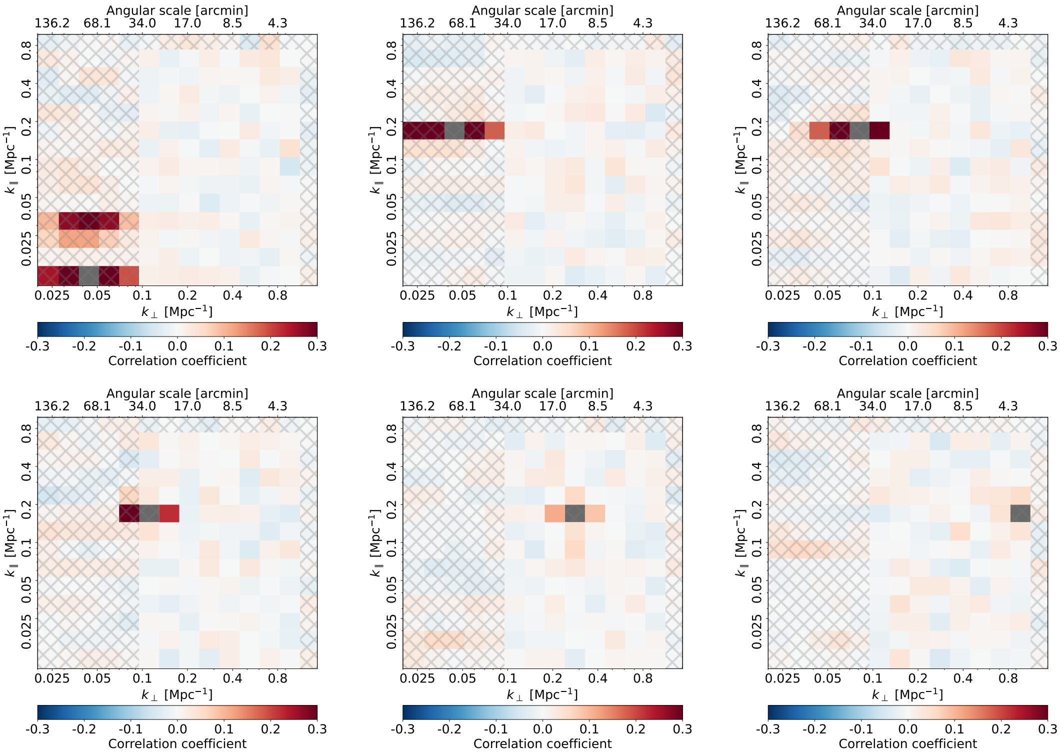


Figure 36: Representative selection of correlation coefficients between the marked gray $\boldsymbol{k}$-bin and all the others. Gray hatching indicates the COMAP S2 mask.

account how the beam of the telescope, the size of the voxels, and the filtering of the data inevitably lead to a degree of signal loss.

The signal loss can mathematically be described by a product of transfer functions so that

$$\langle C(\boldsymbol{k})\rangle = T_{\rm b}(\boldsymbol{k})T_{\rm vox}(\boldsymbol{k})T_{\rm f}(\boldsymbol{k})P_{\rm CO}(\boldsymbol{k}) \equiv T(\boldsymbol{k})P_{\rm CO}(\boldsymbol{k}), \tag{3.35}$$

where the effects of the beam smoothing, the voxel size, and filtering are described in $\boldsymbol{k}$-space by $T_{\rm b}(\boldsymbol{k})$, $T_{\rm vox}(\boldsymbol{k})$ and $T_{\rm f}(\boldsymbol{k})$, respectively.

As described in Sec. 3.4.3, the COMAP mapmaker is based on a simple nearest-neighbor binning scheme. The effect of the binning scheme on the signal is a $\mathrm{sinc}^2(x)$ smoothing in the power spectrum domain. Because the COMAP maps have three dimensions, there will be a corresponding voxel window for each of these. We can write out the voxel transfer function analytically as

$$T_\nu(k_\parallel, k_\perp) = T_{\rm freq}(k_\parallel)T_{\rm pix}(k_\perp) = \mathrm{sinc}^2\left(\frac{\Delta x_\parallel k_\parallel}{2\pi}\right)\mathrm{sinc}^2\left(\frac{\Delta x_\perp k_\perp}{2\pi}\right), \tag{3.36}$$

where the frequency pixel window $T_{\rm freq}(k_\parallel)$ is a function of scales along the line-of-sight $k_\parallel$ and the width of a spectral channel $\Delta x_\parallel$, and analogously for the angular pixel window $T_{\rm freq}(k_\perp)$. Because the angular pixel window is approximately radially symmetric at our scales, we have written $T_{\rm pix}(k_\perp) \approx T_{\rm RA}(k_{\rm RA}) \approx T_{\rm Dec}(k_{\rm Dec})$. The shape of the voxel transfer function can be seen in $(k_\perp, k_\parallel)$-space in Fig. 38, as well as in profile in Fig. 39.

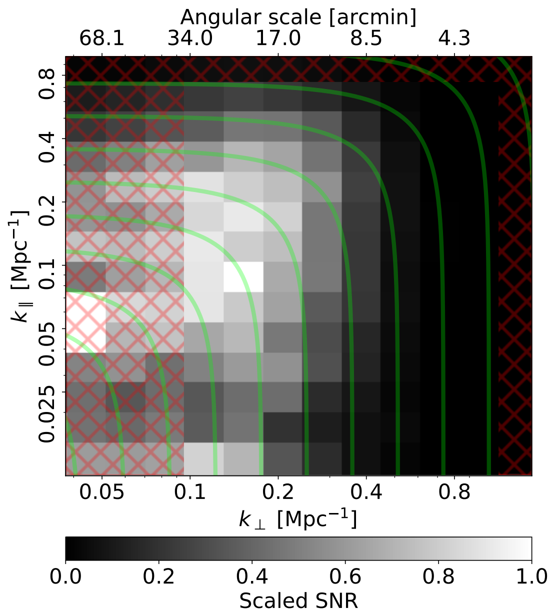


Figure 37: Relative S/N of the COMAP data assuming the Chung et al. (2022a) COMAP fiducial model. Red hatching represents the $\boldsymbol{k}$-space mask applied in the COMAP S2 power spectrum analysis, and green contours represent the bin edges of our 1D spherically averaged power spectra. The S/N normalization is chosen as the maximum value within the uncut region (without hatching) to emphasize the S/N relative to the peak.

Next, the beam transfer function can also be seen in Figs. 38 and 39 (right panel). It is caused by the instrumental beam of the COMAP telescope $B(r)$ (seen in the left of Fig. 39 as a function of radial distance r) smoothing the signal on small scales (see also Sec. 3.5). This effect limits the recovery of small-scale structures the most, as can be seen in Fig. 38. In particular, we can find the effect of the instrumental beam $B(r)$ on the signal in the power spectrum by using the convolution theorem and get

$$T_\mathrm{b}(k_\perp) = |\mathcal{F}\{B(r)\}|^2. \tag{3.37}$$

Additionally, as described in II, the main beam efficiency is applied to the beam prior to computing the transfer function. These (semi)analytical derivations of the beam and voxel window transfer functions are only used in the S2 publications (Papers III-V) as they were deemed more accurate than the simulation-based approach previously used in Early Science (Papers I-II).

Lastly, the pipeline transfer function, seen in Fig. 38, reduces the power of the large-scale modes in our maps. It is somewhat more complicated to estimate as it requires processing simulations through the low-level pipeline described in Sec. 3.4. The exact details of the approach have also changed somewhat between COMAP ES and S2 as the methodology was more refined. However, the rough idea remains similar:

1. Generate a peak-patch dark matter simulation (Bond & Myers, 1996; Stein et al., 2019) and populate the dark matter halos with CO luminosities

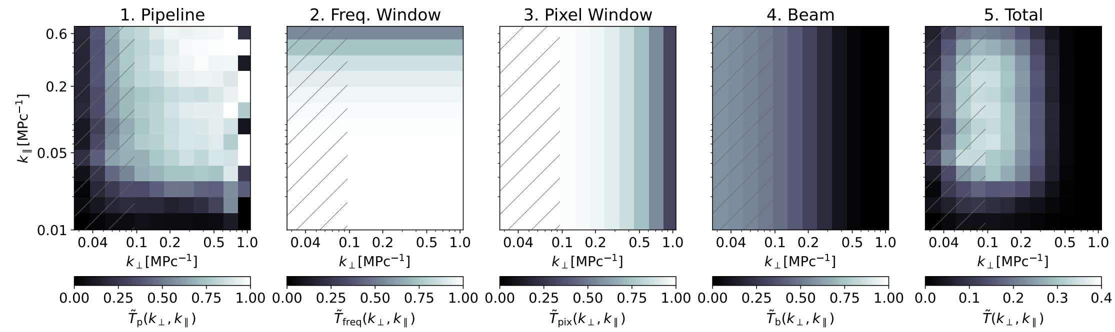


Figure 38: Individual transfer functions quantifying (from left to right) the $\boldsymbol{k}$-space signal loss due to the pipeline filters, the line-of-sight frequency window, the angular pixel window, the finite beam size of the COMAP telescope, as well as all individual effects combined. Scales masked out from the analysis in COMAP S2 are marked by hatching. Courtesy of Jonas Lunde, see Paper III.

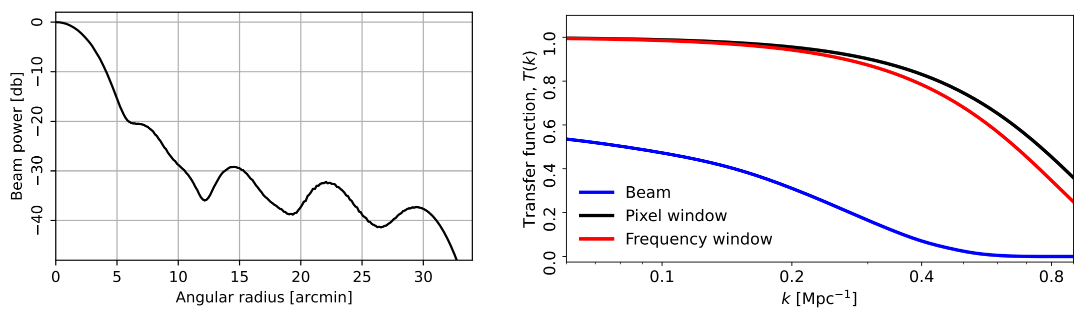


Figure 39: (*Left*) Radially averaged beam profile out to half a degree, and (*right*) the corresponding radial beam transfer function (blue), as well as the frequency and pixel window profiles (red and black). The left and right panels are, respectively, adapted from Papers II and III.

$L_{\rm CO}(M_{\rm halo})$ using the COMAP fiducial model (see 'UM+COLDz+COPSS' by Chung et al., 2022a) (or any other model).

2. Optionally boost the signal to create a detectable signal at the chosen S/N.

3. Inject the signal into the raw COMAP data, which serves as the noise term in the data model. Given the telescope (bi-linearly interpolated) pointing P_L, gain G, and beam B we can write the resulting mock TOD as

$$d_{t,\nu}^{\rm mock} = \mathsf{GR}(\theta_0)\mathsf{P}_L\mathsf{B}s_{\theta,\nu}^{\rm mock} + n_{t,\nu}. \tag{3.38}$$

Because the simulated signal cubes $s_{\theta,\nu}^{\rm mock}$ are generated as if they were in a Cartesian grid, i.e., a small patch along the celestial equator (zero declination), we must rotate the telescope pointing P_L to be centered on the equator using an Euler rotation $\mathsf{R}(\theta_0)$. This is done to ensure the injected signal inherits all the correct geometric distortion from the Plate Carrée projection used in the maps. In Fig. 40 we show examples of some signal-injected TODs.

4. After constructing the mock TODs $d^{\text{mock}}_{t,\nu}$ we run it though the pipeline as described in Paper III and Sec. 3.4. We can write this as
$$\hat{\boldsymbol{m}}^{\text{mock}}_{\theta,\nu} = f_{\text{map}}\left[M(f_{\text{TOD}}[\mathsf{GR}(\theta_0)\mathsf{P}_L\mathsf{B}s^{\text{mock}}_{\theta,\nu} + n_{t,\nu}]\right], \tag{3.39}$$
with the time- and map-domain filters in the pipeline being denoted as f_{TOD} and f_{map}, respectively, while the mapmaker is denoted by M. Meanwhile, the signal TOD used in Eq. (3.38) is binned up into maps without any filtering so that
$$\hat{\boldsymbol{s}}^{\text{mock}}_{\nu,\theta} = M(\mathsf{GR}(\theta_0)\mathsf{P}_L\mathsf{B}s^{\text{mock}}_{\nu,\theta}) \tag{3.40}$$
Figure 41 shows some examples of signal-injected and filtered maps and the unfiltered signal.

5. The final pipeline transfer function can then be estimated using a cross-power spectrum (as described in Eq. (3.31)) between the mock map $\hat{\boldsymbol{m}}_{\text{mock}}$ and unfiltered signal $\hat{\boldsymbol{s}}^{\text{mock}}_{\nu,\theta}$, normalized to the auto-power spectrum of the signal $\hat{\boldsymbol{s}}^{\text{mock}}_{\nu,\theta}$ so that
$$\hat{T}_f \approx \frac{C(\hat{\boldsymbol{m}}_{\text{mock}}, \hat{\boldsymbol{s}}^{\text{mock}})}{P(\hat{\boldsymbol{s}}^{\text{mock}})}. \tag{3.41}$$
This estimator will then quantify how much the filtered and unfiltered signals have in common. At the same time, by using a cross-spectrum, we ensure that the transfer function estimator is robust against any residual systematics present in the raw data $n_{\nu,\theta}$ in Eq. (3.38). Meanwhile, in Paper I we used a less optimal estimator that was purely based on auto-power spectra, as well as having a noise bias subtraction step:
$$\hat{T}_f \approx \frac{P(\hat{\boldsymbol{m}}_{\text{mock}}) - P(\hat{\boldsymbol{m}}_n)}{P(\hat{\boldsymbol{s}}^{\text{mock}})}, \tag{3.42}$$
where $\hat{\boldsymbol{m}}_n$ is the map of the raw COMAP data without added signal.[5]

With the individual transfer functions in hand, the full transfer function T can deconvolved from the FGPXS signal estimator so that
$$P_{\text{CO}}(\boldsymbol{k}) = \frac{C(\boldsymbol{k})}{T(\boldsymbol{k})}, \tag{3.43}$$
with the corresponding uncertainties
$$\sigma^{\text{CO}}_{P(\boldsymbol{k})} = \frac{\sigma_{C(\boldsymbol{k})}}{T(\boldsymbol{k})}. \tag{3.44}$$
The effect of the transfer function on the error bars can be seen in Fig. 43. We can see that the uncertainties of the COMAP data points increase at the lowest and highest k where, respectively, the filter and beam transfer functions remove most of the structure in the map.

[5] In addition, the Paper I transfer function estimator used a regridded signal simulation to estimate $\hat{\boldsymbol{s}}^{\text{mock}}$, instead of binning the injected unfiltered signal TOD as in the 4th step. This introduced an implicit pixel window difference because the injected and reprojected signals did not always end up in the same output pixels. The result was a spurious signal loss estimate at small perpendicular scales. However, this was understood and fixed in the COMAP S2 release of Paper III.

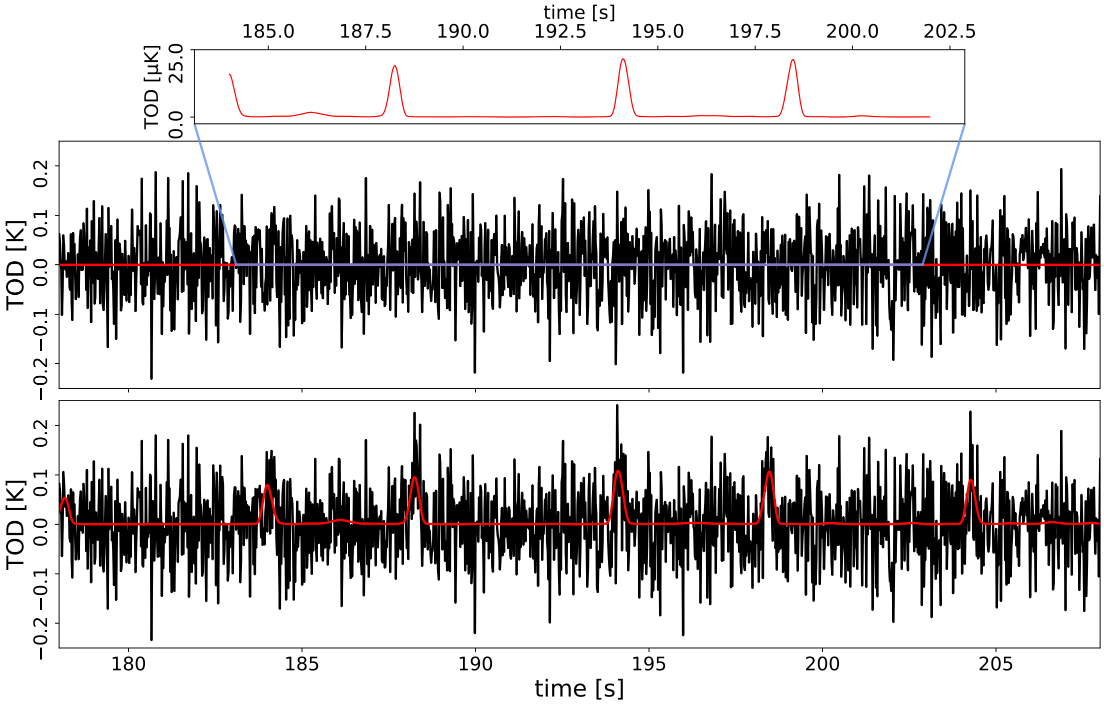


Figure 40: Example of signal-injected TODs in a (*top*) noise and (*bottom*) signal dominated case. The signal strength corresponds to one and five thousand times the Chung et al. (2022a) COMAP fiducial model, respectively.

3.6.4 Power spectrum uncertainties and null tests

Before wrapping up Sec. 3.6 with an overview of the COMAP CO(1–0) power spectrum constraints and modeling, we will mention how the power spectrum uncertainties are typically estimated. We will also discuss how to perform null tests on the data to ensure that any residual systematic effects present are below the noise level. Specifically, developing a more accurate data-driven uncertainty estimate and an improved null test framework for the COMAP S2 data release is one of the main contributions to this thesis. We will thus only skim the uncertainty and null test frameworks of Paper II and go into somewhat more detail on the methodologies of Paper IV in this section.

Uncertainty estimation

When estimating uncertainties for power spectra, there are generally two approaches: simulations and data-driven methods. The first approach, chosen in Paper II for its simplicity, is noise simulations. The basic idea is to generate an ensemble of maps with only noise, compute power spectra from these, and estimate uncertainties based on these. The noise simulations should have the same properties as the actual data. Hence, the approach can be made arbitrarily complicated, depending on how realistic the noise model is.

In the first season of COMAP releases (Paper II) a relatively simple approach was chosen, where white noise realizations $\boldsymbol{m}^i_{\nu,\theta}$ are drawn from the voxel

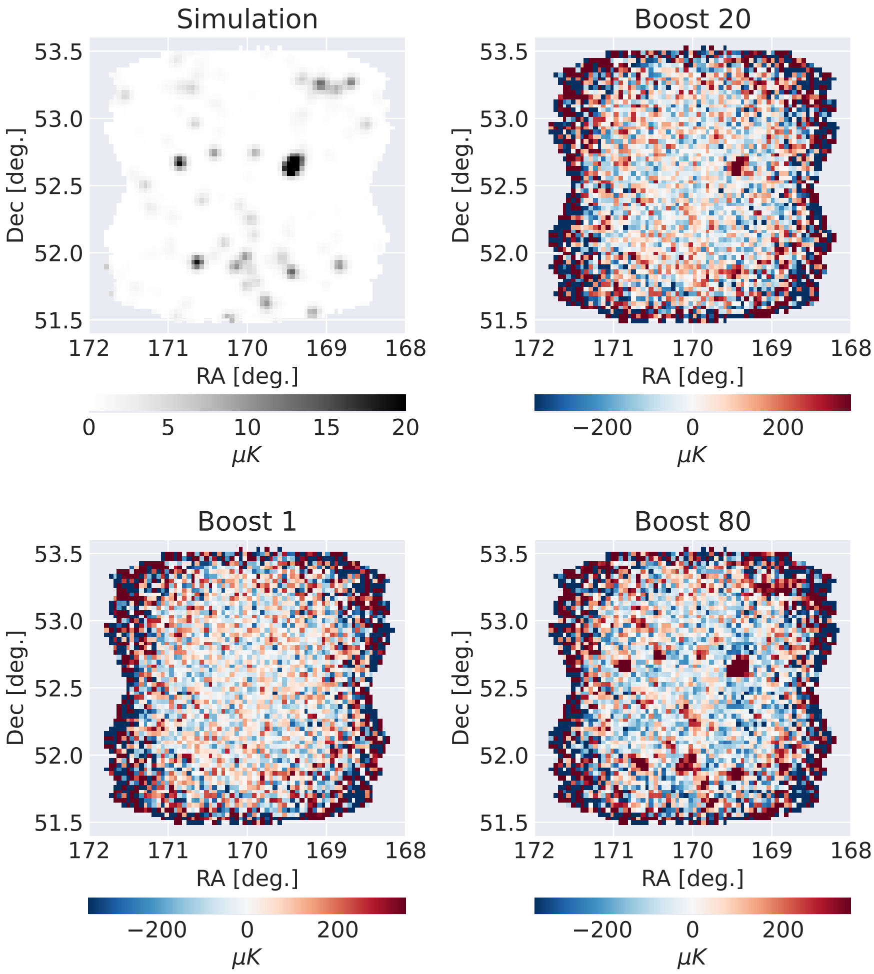


Figure 41: Example of unfiltered signal (top left) and several map frequency slices with signal-injected and filtered COMAP data corresponding to different signal strengths (boosts from fiducial strengths). The signals are produced using the Chung et al. (2022a) COMAP fiducial model. Courtesy of Jonas Lunde, see Paper III.

uncertainty map $\boldsymbol{\sigma}_{\boldsymbol{m}_{\nu,\theta}}$ so that

$$\boldsymbol{m}^i_{\nu,\theta} \sim \mathcal{N}(0, \boldsymbol{\sigma}_{\boldsymbol{m}_{\nu,\theta}}). \tag{3.45}$$

From a large number of independent white noise realizations, an ensemble of power spectra is generated from which one can estimate the power spectrum uncertainty $\sigma_{C(\boldsymbol{k})}$.

The advantage of using this method is that it is very quick to run due to its simplicity and generates fairly accurate uncertainty estimates on small scales where the noise properties of the map are closest to white noise. However, as we can see from the transfer function estimate in Fig. 38, the pipeline transfer function results in a loss of structure on large scales in the maps. Additionally, there is some residual correlated noise in the data (see Fig. 9. of Paper III). Therefore, the result of using this approach is that the noise estimate will overestimate the amount of noise in the data on large scales due to low-level filtering. Meanwhile, on small scales, the noise will be underestimated due to residual correlated noise that is not accounted for. This is seen in Fig. 32 in the bottom right panel, showing the 2D field coadded FPXS in units of the estimated uncertainty.

Due to COMAP S2's dramatically increased data volume, the simple white noise simulation-based approach was found to no longer be accurate enough. At the same time, there is not yet an existing framework in which realistic TOD simulations can be propagated all the way through the COMAP pipeline to model the uncertainty appropriately.

Therefore, a data-driven approach was chosen for S2 (Paper IV), where all computed uncertainties are directly estimated from the data. As such, they inherit all filtering biases and noise properties of the real data. Specifically, where the ES approach estimated power spectrum uncertainties from a large number of noise simulations, the S2 approach generated a large ensemble of randomized null difference (RND) maps

$$\Delta \boldsymbol{m}_i^{\mathrm{RND}} = \frac{\boldsymbol{m}_{A,i}^{\mathrm{RND}} - \boldsymbol{m}_{B,i}^{\mathrm{RND}}}{2}, \tag{3.46}$$

where $\boldsymbol{m}_{A,i}^{\mathrm{RND}}$ and $\boldsymbol{m}_{B,i}^{\mathrm{RND}}$ maps are made from randomly shuffled halves of all the scans in the data. The resulting RND maps $\Delta \boldsymbol{m}_i^{\mathrm{RND}}$ then have the same noise properties as the actual data but should not contain signal or coherent systematic effects because of the difference between two random halves. From the $\Delta \boldsymbol{m}_i^{\mathrm{RND}}$ maps, we simply compute uncertainties the same ways as from the noise simulation maps. The resulting error bars are used, for example, in Fig. 33, whereas the older noise simulations are used in Fig. 32. To judge the quality of the estimated errors, we can compare the COMAP S2 FGPXS in units of its uncertainties (in the lower right of Fig. 33)) with the equivalent FPXS of COMAP ES (Fig. 32). We see that the noise appears to be much more evenly distributed in $\boldsymbol{k}$-space in the COMAP S2 FGPXS than for the COMAP ES FPXS. Additionally, we no longer see the same under- or overestimation of noise on, respectively, large and small scales. We can, therefore, conclude that the uncertainty estimates capture the properties of the data more accurately.

Null tests

Lastly, to ensure that the power spectra (or any other summary statistic) behave according to the expectations of the experimental noise and do not contain significant systematic residuals, we need to perform a set of null tests. The null test framework was one of the things that changed significantly between the two seasons of COMAP (i.e., Papers II and IV). The null test framework was essentially redesigned from scratch for Paper IV.

In Paper II, six null tests are performed. Each computes an FPXS (see Sec. 3.6.2) where the two cross-correlated maps are of different fields. By doing so, we test the null hypothesis that independent fields at different elevations should be consistent with noise if there are no common systematic effects in the two cross-correlated fields. Because the extragalactic CO in independent fields (without any sensitivity for a monopole due to our normalization filter) is statistically independent, there is no CO signal contribution to cross-field spectra. However,

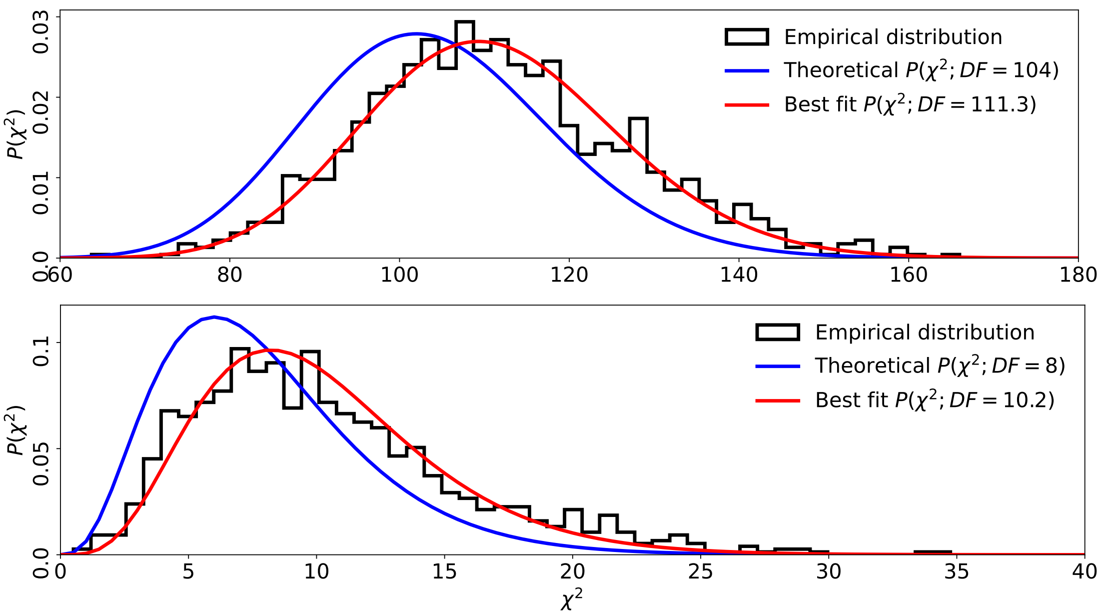


Figure 42: Empirical χ^2 distribution (black) of the cylindrically and spherically averaged FGPXS (*top* and *bottom*, respectively), the corresponding theoretical χ^2 distribution (blue) assuming degrees-of-freedoms (DF) equal to the number of summed $\boldsymbol{k}$-bins, and χ^2 distributions with a variable DF fit to the empirical distribution (red). The theoretical χ^2 distribution is, in both cases, a poor match for the empirical distribution, as indicated by a clear shift between the black and blue curves.

the problem with using cross-spectra between fields is that it can be challenging to find a meaningful common coordinate system to rotate both fields to before cross-correlation. Additionally, it only allows for null tests in which two maps are crossed and not tests in which only one field is independently tested by itself. Lastly, it is somewhat challenging to perform a high number of null tests in this framework.

To improve this for the order of magnitude higher degree of sensitivity of the S2 power spectra, we developed a new difference map-based test scheme in Paper IV. Subsequently, half-difference maps, just like in Eq. (3.46), but with meaningful non-randomized splits, are computed. Next, FGPXS of the difference maps are estimated just like for regular maps (see Sec. 3.6.2). Thus, the null hypothesis of our COMAP S2 null tests is the following: In computing the difference between splits, the resulting map should be void of signal, while any systematic not shared across splits should remain. Hence, the FGPXS of the difference maps should be consistent with noise expectations.

Specifically, a set of 26 different variables (related to the weather, telescope pointing, pipeline, and noise parameters; see Table C.2 in Paper IV) was identified, and the data are evenly split into two for each of these variables. The RND uncertainty estimation is computationally heavy, and we currently do not support generating RND ensembles for each individual null test split map. We, therefore, make sure that each null split is evenly distributed across elevation as well as when the fields are rising and setting. This enables us to use the same RND-derived uncertainties for all 26 null test variables using the same RND ensemble. However, the disadvantage of doing this is that the sensitivity of our null tests is somewhat reduced. Generalizing and optimizing the framework to uneven splits will be a point of future improvement.

In addition to the 26 null test variables over which a difference map is computed, we independently perform each null test in the three Fields. We also individually process scans with a slow and fast azimuth speed, as they could have different mechanical vibration issues. The FGPXS computed for each difference map is averaged both spherically and cylindrically. From this, we get 312 independent null tests in total that we can perform to judge the data quality of the final power spectra of S2.

To say whether the null test FGPXS are statistically consistent with noise, we perform a χ^2 test. For each null test variable j, we compute

$$\chi^2_{\mathrm{null},j} = \sum_{\mathbf{k}_i} \left(\frac{C^{\mathbf{k}_i}_{\Delta m_j} - \mu^{\mathbf{k}_i}_{\Delta m_j}}{\sigma_{C^{\mathbf{k}_i}_{\Delta m_j}}} \right)^2 = \sum_{\mathbf{k}_i} \left(\frac{C^{\mathbf{k}_i}_{\Delta m_j}}{\sigma_{C^{\mathbf{k}_i}_{\Delta m_j}}} \right)^2 , \tag{3.47}$$

where $\mu^{\mathbf{k}_i}_{\Delta m_j} = 0$ according to the null hypothesis. The difference map FGPXS $C^{\mathbf{k}_i}_{\Delta m_j}$ has (a RND-estimated) uncertainty $\sigma_{C^{\mathbf{k}_i}_{\Delta m_j}}$ in bin $\mathbf{k}_i$. From this, we can obtain a probability-to-exceed (PTE) by comparing the resulting $\chi^2_{\mathrm{null},j}$ to their expected distribution as

$$\mathrm{PTE}(\chi^2) = 1 - \mathrm{CDF}(\chi^2), \tag{3.48}$$

where $\mathrm{CDF}(\chi^2)$ denotes the cumulative probability function of the $\chi^2_{\mathrm{null},j}$.

However, we cannot simply assume that the $\chi^2_{\mathrm{null},j}$ will adhere to a simple theoretical χ^2 distribution. It turns out that the cross-spectrum between two independent Gaussian random maps is itself not Gaussian random but rather distributed according to a modified Bessel function (Watts et al., 2020; Nadarajah & Pogány, 2016; Gaunt, 2019). But as we additionally bin our FGPXS estimates (either spherically or cylindrically), the noise properties in each bin will again be driven toward Gaussian properties to some extent by the central limit theorem. Therefore, there is no analytically well-defined expression for the distribution of our empirical $\chi^2_{\mathrm{null},j}$ values.

Instead of relying on theoretical expressions, we can use the RND methodology, otherwise used for estimating power spectrum uncertainties. As we saw earlier, each RND map is not different from a proper null test difference map except that the data are split randomly. All systematics and CO signals should cancel in an RND map and, therefore, fulfill the null hypothesis perfectly. As such, they are an excellent way to empirically test whether a proper null test difference map behaves according to noise expectations of the data or if there are remaining systematic errors.

In Fig. 42, we show an example of the empirical $\chi^2_{\mathrm{null},j}$ distribution for both the cylindrically and spherically averaged (masked) FGPXS estimated from hundreds of RND spectra. We see that the theoretical χ^2 distribution with degrees-of-freedom (DF) equal to the number of summed $\boldsymbol{k}$-bins underestimates the empirical

χ^2 values. This would result in the null test spuriously failing as the empirical distribution is shifted to the high tail of the χ^2 distribution. To provide a slightly better fit to the empirical distribution, we fit a χ^2 distribution with variable degrees-of-freedom to the empirical distribution. We see from Fig. 42 that this modified χ^2 distribution provides a much better fit to the empirical distribution than if a fixed DF was used. The fit distribution can be used to evaluate Eq. (3.48) in the far tails when evaluating the PTE for outliers.

Finally, the PTEs of the null tests are expected to be drawn from a uniform distribution. To validate this, we can perform a Kolmogorov–Smirnov (KS) test to see if the null test PTE ensemble is plausibly drawn from a uniform distribution.

As mentioned, we get a total number of 312 null tests (see Appendix C of Paper IV for a full table of PTEs). These behave according to noise expectations, and the resulting PTEs all pass the KS uniformity test, as discussed in Paper IV. We can thus confidently say that the final data products of the COMAP S2 release seen in Figs. 43 and 33 are free from systematic effects down to the instrumental noise level.

3.6.5 The state-of-the-art CO LIM constraints and connecting them to physics

The state-of-the-art CO LIM constraints

Putting it all together, the S2 power spectra of Paper IV provide a dramatic improvement in sensitivity compared to the Season 1 constraints, despite the slightly less sensitive FGPXS estimator and large-scale $k_\perp$-cuts. The improvement is easily seen in Fig. 43, the main result from Paper IV. The COMAP S2 points show around an order of magnitude improvement in the error bar size compared to the COMAP ES power spectrum (Paper II) and the COPSS measurement of Keating et al. (2016). The updated COMAP power spectrum data now scatter much tighter around the two brightest non-excluded models, the COMAP fiducial model of Chung et al. (2022a) and the Li-Keating model (Keating et al., 2020).

As such, the COMAP S2 power spectrum is the currently tightest direct 3D LIM constraint on the CO(1–0) clustering power spectrum in the literature. However, the uncertainties are still too large to claim any detections, let alone separate models. Nevertheless, we note that we have a 2.7σ excess in the second power spectrum bin in Fig. 43. At 2.7σ the significance of the excess is not yet enough to claim any detection, as it could simply be a noise fluctuation or some systematic effect that has not yet been discovered. Even still, it is encouraging to see an excess in our FGPXS after passing all null tests, as this could possibly also represent a first subtle hint of cosmic CO in the COMAP data if it turns out not to be explained by alternative causes.

Compared to COPSS, the only comparable CO(1–0) LIM experiment to COMAP, we see from Fig. 43 that there is a mild 2.5σ tension between the sixth COMAP bin t $k \sim 0.6\,\mathrm{Mpc}^{-1}$ and one of the 2.5σ excess points of COPSS. However, we do not consider this tension statistically significant at this point, and this point represents the only noteworthy disagreement between the two surveys.

For more details and discussion on the COMAP ES and S2 power spectrum results, we refer the interested reader to Papers II and IV.

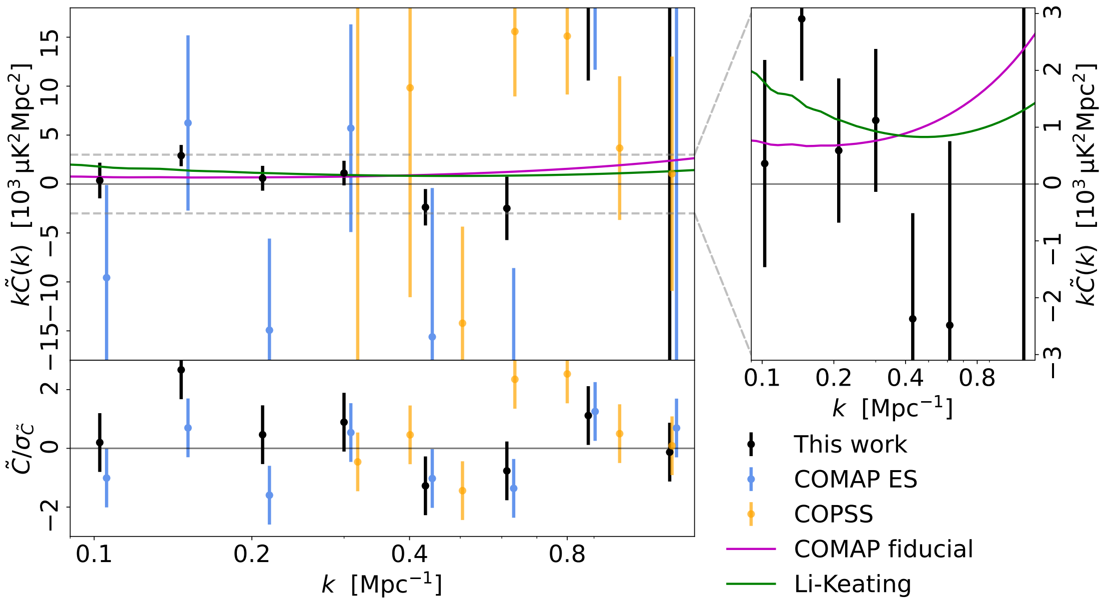


Figure 43: (*Upper panel*) Combined CO power spectrum estimates with 1σ uncertainties, using only CES data and all three fields, from COMAP ES (blue) and COMAP S2 (black) in relation to the power spectrum points of COPSS (Keating et al., 2016) (orange) and the two closest CO(1–0) power spectrum models; the COMAP fiducial (Chung et al., 2022a) and Li-Keating (Keating et al., 2020) models. (*Lower panel*) Same data points as in the upper panel, but in units of their 1σ uncertainty. (*Inset*) Zoom in on COMAP season 2 data points and the two selected models. The figure is adapted from Paper IV.

From power spectra to physics

Up until this point, we have only considered how to filter, calibrate, and compute power spectra as a summary statistic from the COMAP data. However, as astrophysicists and cosmologists, we are also interested in what the data can tell us about the physics of the EoGA (and eventually EoR). To do so, we need to first write down a model of how the physics, like the star formation rate (SFR), connects to our observable. In our case, the observable is the CO(1–0) power spectrum. This is the topic of Paper V. In Paper V, we take the COMAP S2 results produced by the power spectrum analysis of Paper IV and infer astrophysical parameters from it.

Although most of the work in this thesis has focused on the lower-level aspects of COMAP and less so on the work presented in Paper V, we still include a summary of the modeling as it is important to understand the ultimate goals and the context of COMAP as a whole. For more details, we therefore refer the interested reader to Paper V, Chung et al. (2022a), and (Breysse et al., 2022), as well as references therein.

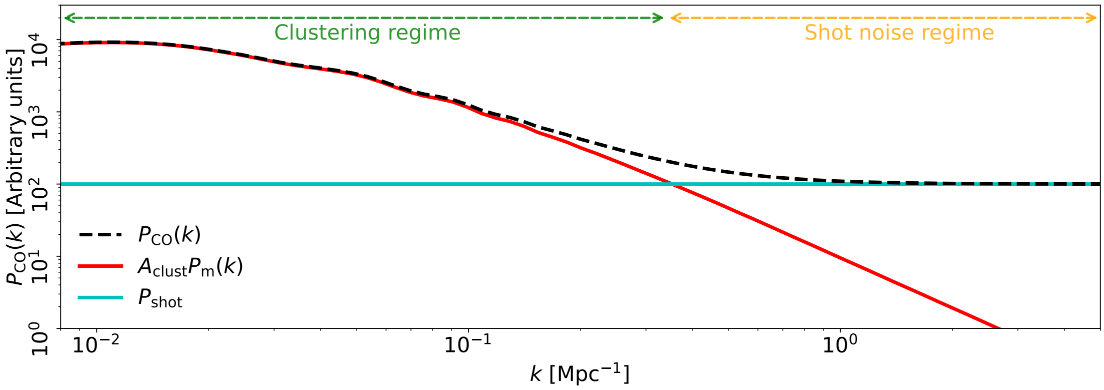


Figure 44: Typical CO line intensity mapping power spectrum $P_{\rm CO}(k)$ (black) made up of a clustering term $A_{\rm clust}P_{\rm m}(k)$ (red) and a constant shot noise contribution $P_{\rm shot}$ (blue), in arbitrary units. Note that this plot is plotted in units without tilt (i.e., $P(k)$) as opposed to the power spectra in Figs. 43 and 46 (with $kP(k)$).

We start by writing down an expression for a typical line intensity power spectrum as

$$P_{\rm CO}(k) = A_{\rm clust}P_{\rm m}(k) + P_{\rm shot}. \tag{3.49}$$

Because the star-forming CO emission field traces the underlying dark matter structures on large scales, the first term called the clustering spectrum, is given by the dark matter power spectrum $P_{\rm m}(k)$ times a scaling factor $A_{\rm clust}$. This quantifies how tightly the luminous matter tracks the underlying dark matter. However, CO emitters are also discrete sources. There is, therefore, a constant Poisson noise term, called shot noise $P_{\rm shot}$, which dominates the emission on small scales. In Fig. 44 an example of this decomposition is shown.

If we were to measure the CO field directly in comoving coordinates, we could write the clustering amplitude $A_{\rm clust} = \langle Tb\rangle^2$, i.e., the mean line temperature-bias product. This mean line temperature-bias product can itself be written as

$$\langle Tb\rangle \propto \int \mathrm{d}M_{\rm h}\,\frac{\mathrm{d}n}{\mathrm{d}M_{\rm h}}L(M_{\rm h})b_{\rm h}(M_{\rm h}), \tag{3.50}$$

where $M_{\rm h}$ is the virial mass of a halo, $L(M_{\rm h})$ is the CO luminosity, $\frac{\mathrm{d}n}{\mathrm{d}M_{\rm h}}$ is the differential halo mass function, and $b_{\rm h}(M_{\rm h})$ is the bias that quantifies how tightly a halo traces the continuous dark matter density field. Furthermore, we can expand the shot noise to be written as

$$P_{\rm shot} \propto \int \mathrm{d}M_{\rm h}\,\frac{\mathrm{d}n}{\mathrm{d}M_{\rm h}}L^2(M_{\rm h})b_{\rm h}(M_{\rm h}). \tag{3.51}$$

As such, Eqs. (3.51) and (3.50), respectively, quantify the first and second moments of the CO luminosity function and contain information about how the CO structures trace the underlying dark matter, as well as the discrete nature of the emitters that make up the CO emission (Kovetz et al., 2017).

However, a LIM map does not directly measure the structures in comoving coordinates but instead in redshift space (i.e., the line-of-sight distance to some

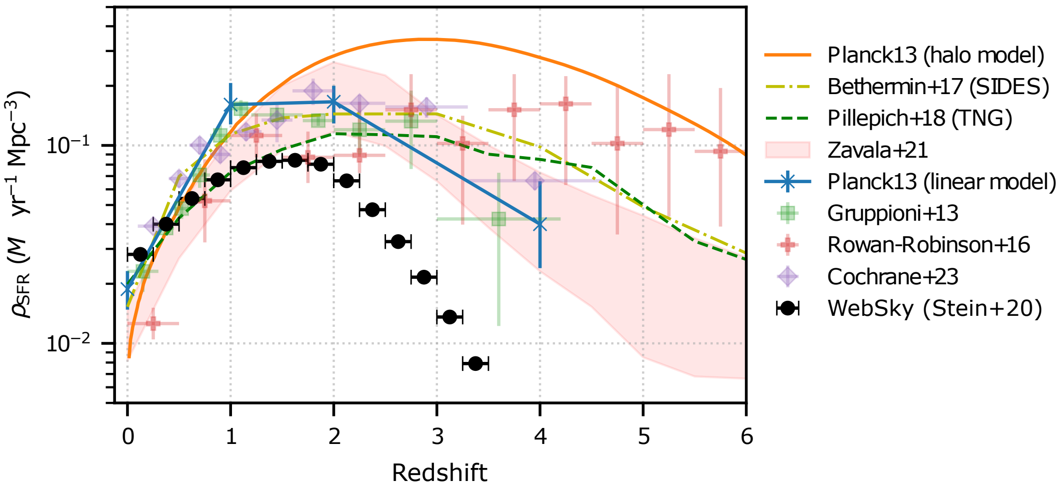


Figure 45: Star formation rate density, $\rho_{\rm SFR}$, estimates of different surveys from low to high redshift (Béthermin et al., 2017; Pillepich et al., 2018; Zavala et al., 2021; Gruppioni et al., 2013; Rowan-Robinson et al., 2016; Cochrane et al., 2023; Stein et al., 2020; Planck Collaboration XXX., 2014). Courtesy of Dongwoo Chung (private communication).

object is inferred through the cosmic redshift z; see, e.g., Schneider, 2015). The structures in the map are, therefore, affected by redshift space distortions (RSD) that can lead to some enhancement or attenuation in the CO power spectrum. One of these effects is the Kaiser effect (Kaiser, 1987; Hamilton, 1998), where the large-scale structure in the map will increase clustering due to infalling structures. On the other hand, there is a loss of small-scale structure along the line-of-sight due to line broadening of the CO line (Chung et al., 2021). This has been estimated to reduce the power spectrum of CO structures at scales 0.2–0.3 Mpc^{-1} by around $\sim 10\,\%$ (Chung et al., 2021).

While the shape of the power spectrum can be affected to a degree by RSDs, the overall amplitude of the CO power spectrum can vary by orders of magnitude depending on assumptions of the physics of the CO emitters on high redshift. Especially the assumptions of halo and galaxy properties, like star formation, at high redshift and how these affect the CO observed CO luminosity have a huge impact on the magnitude of the CO power spectrum. For example, in Fig. 45, we can see the cosmic star formation density $\rho_{\rm SFR}$ as a function of redshift. At the COMAP-Pathfinder redshift range $z \sim 2$–3 the estimates of $\rho_{\rm SFR}$ already span 0.5–1 dex. As a result, predictions for the CO(1–0) power spectrum at the EoGA vary by orders of magnitudes depending on their model and SFR assumptions at early times. This spread in model predictions can be seen in Fig. 46 (we refer the interested reader to Pullen et al., 2013; Li et al., 2016; Padmanabhan, 2018; Chung et al., 2022a; Keating et al., 2020; Yang et al., 2022, for details on the modeling landscape).

This is exactly where COMAP will significantly improve our understanding of EoGA star formation physics by pushing down the upper limits on the CO power spectrum until a detection is made. In fact, with the results from Paper II, COMAP ES was already able to exclude both the Padmanabhan (2018) model with

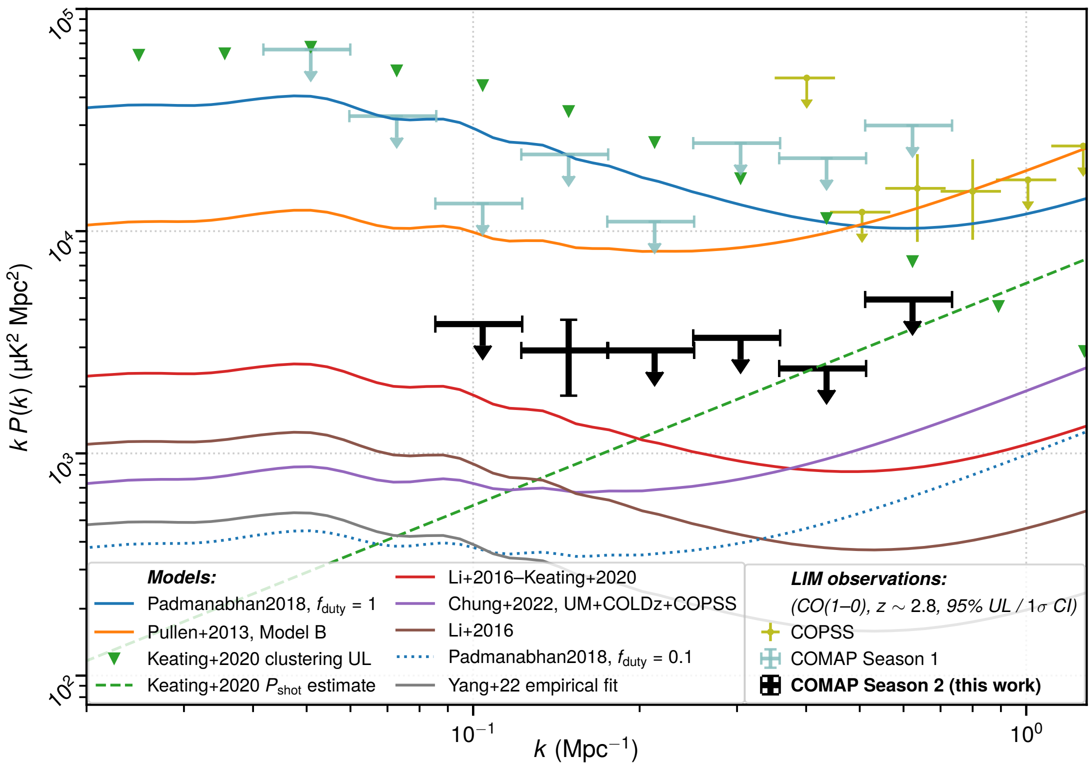


Figure 46: Landscape of CO(1–0) power spectrum models, with the upper limits at 95 % confidence of COPSS (Keating et al., 2016) and COMAP ES (derived from the ES data points of Paper II), and COMAP S2 (Paper IV) per k-bin. Courtesy of Dongwoo Chung, see Paper V.

$f_{\text{duty}} = 1$ and Model B by Pullen et al. (2013). However, using our significantly improved sensitivity of the COMAP S2 power spectra from Paper IV, we can further exclude the two models in *individual* k-bins. The COMAP sensitivity is expected to increase further as more data are collected and the lower levels of the analysis improve. Thus, upper limits are expected to migrate further towards where most models predict a first detection within the next couple of years.

Furthermore, the space of models can be divided into two categories: They either directly model the dark-matter-halo-to-CO connection with some constraints from observations of the CO luminosity function (the Padmanabhan 2018 and the COMAP fiducial model by Chung et al. 2022a), or tie together halo properties and CO luminosity via some intermediate proxy like the star formation rate or infrared luminosity (such as Pullen et al., 2013; Li et al., 2016; Keating et al., 2020; Yang et al., 2022).

As an example of one of these models, we consider the COMAP fiducial model by Chung et al. (2022a), in which $L(M_\text{h})$ is modeled as a double power law in combination with priors from the CO Luminosity Density at High-z (COLDz) survey (Pavesi et al., 2018; Riechers et al., 2019) and UniverseMachine (UM) (Behroozi et al., 2019). It is parameterized as

$$\frac{L'_{\text{CO}}(M_\text{h})}{\text{K}\,\text{km}\,\text{s}^{-1}\,\text{pc}^2} = \frac{C}{(M_\text{h}/M)^A + (M_\text{h}/M)^B}, \tag{3.52}$$

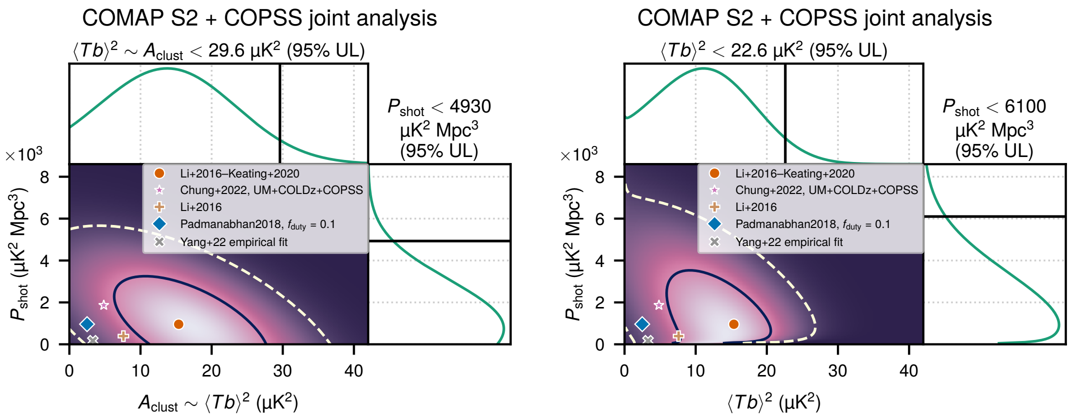


Figure 47: Posterior contours (black solid and white dashed 2D contours mark 39 % and 86 % probability, and the solid horizontal line shows the 95 % upper limit on the marginal probability distributions) for the clustering and shot noise amplitudes of the CO power spectrum, assuming both a (*left*) "b- and v_{eff}-agnostic" and (*right*) "b- and v_{eff}-informed" two-parameter model. The data used is the COMAP S2 data from Paper IV combined with COPSS data (Keating et al., 2016). The clustering and shot noise amplitudes of a selection of models from Fig. 46 are included for reference. Courtesy of Dongwoo Chung, see Paper V.

and

$$\frac{L_{\mathrm{CO}}(M_{\mathrm{h}})}{L_\odot} = 4.9 \cdot 10^5 \frac{L'_{\mathrm{CO}}(M_{\mathrm{h}})}{\mathrm{K\,km\,s^{-1}\,pc^2}}, \tag{3.53}$$

with informative UM priors on the four (dimensionless) parameters

$$A = -1.66 \pm 2.33, \tag{3.54}$$
$$B = 0.04 \pm 1.26, \tag{3.55}$$
$$\log C = 10.25 \pm 5.29, \tag{3.56}$$
$$\log(M/M_\odot) = 12.41 \pm 1.77, \tag{3.57}$$

governing the average relation and a log-normal scatter added about the average relation with an initial prior of $\sigma = (0.4 \pm 0.2)$ dex similar to Li et al. (2016). The other models are similarly parametrized, with possible intermediate steps for the halo-CO relations.

Using what we have defined above, we can, as outlined in more detail in Paper V, perform several analyses that may give us insight into the star formation at early times. As a starting point, we can perform a simple two-parameter model in the form of Eq. (3.49), with a clustering amplitude and shot noise level A_{clust} and P_{shot} as free parameters. Furthermore, we can either be agnostic about assumptions of specific models and RSD or assume specific values for the RSD. These are referred to as the "b- and v_{eff}-agnostic" and "b- and v_{eff}-informed" two-parameter analyses in Paper V. This type of analysis will not provide details on the actual physics inside halos. Still, it may provide estimates of the amount of clustering and shot noise in the system and, as such, constrain some aspects of the luminosity function $L(M_{\mathrm{h}})$. To obtain more detailed pictures of the physics one needs to assume a model like the COMAP fiducial model (Chung et al., 2022a), or one of the other models, which focuses more on (empirical) physical relations. This is referred to as the "five-parameter" in Paper V. Subsequently, one can map out the posterior distribution of the model space using some standard Markov Chain Monte Carlo (MCMC) algorithm and obtain parameter estimates with associated uncertainties and correlations.

As we can see in Fig. 47 for both the informed and agnostic models, the two-parameter posterior probability analysis prefers the Li-Keating model Keating et al. (2020), but the sensitivity of COMAP is not yet large enough to claim any detection. However, the other (non-excluded) models are still consistent with the data at 2σ in both analysis modes. Interestingly, from the five-parameter analysis, we see the first hints of an increase in the faint end of the CO luminosity function, as well as in clustering, as shown in Fig. 48. For more details and discussion of the COMAP modeling results, see Paper V, and the works of Chung et al. (2022a).

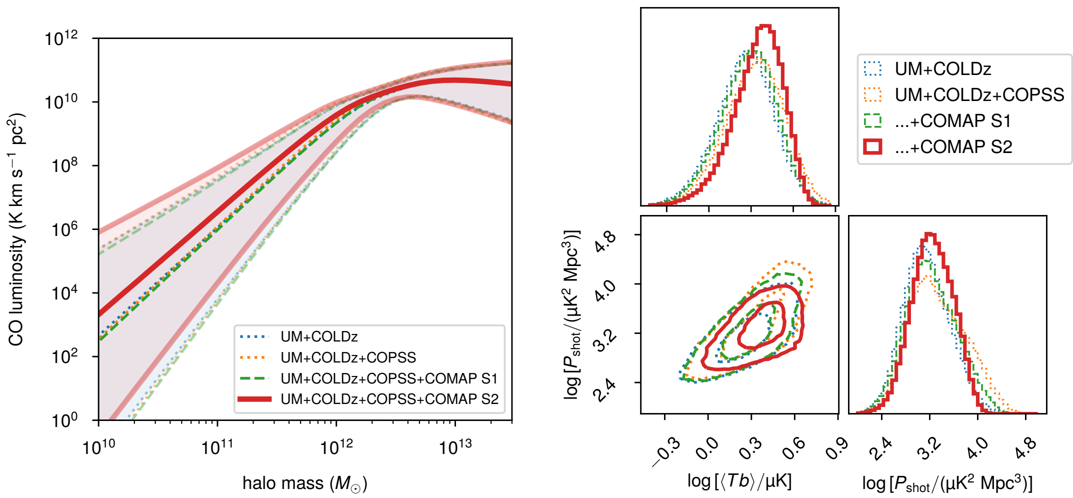


Figure 48: *Left*: Median and 68 % confidence of the luminosity function $L'_{\rm CO}$ from Eq. (3.52) assuming a five-parameter COMAP fiducial model of Chung et al. (2022a). The different curves show the "UM+COLDz" priors, as well as when including in COPSS Keating et al. (2016), COMAP ES, and COMAP S2 data from Papers II and IV. *Right*: Posterior probability (1σ and 2σ) contours on $\langle Tb\rangle$ and $P_{\rm shot}$ resulting from the five-parameter analysis. Courtesy of Dongwoo Chung, see Paper V.

3.7 Detection of CMB and continuum point sources

Now that we have introduced the COMAP experiment and explained how we have contributed to and significantly improved its data analysis, we will briefly present an interesting discovery made in the COMAP maps. When looking at the frequency-averaged COMAP maps, it turns out that they contain a residual of the CMB and a row of point sources. As these results have not yet been published elsewhere, we present the continuum residuals in this work.

In Fig. 49, we show an example of the CMB detection in COMAP Field 2 and the corresponding Planck LFI 30 GHz and Planck 2018 Commander CMB maps

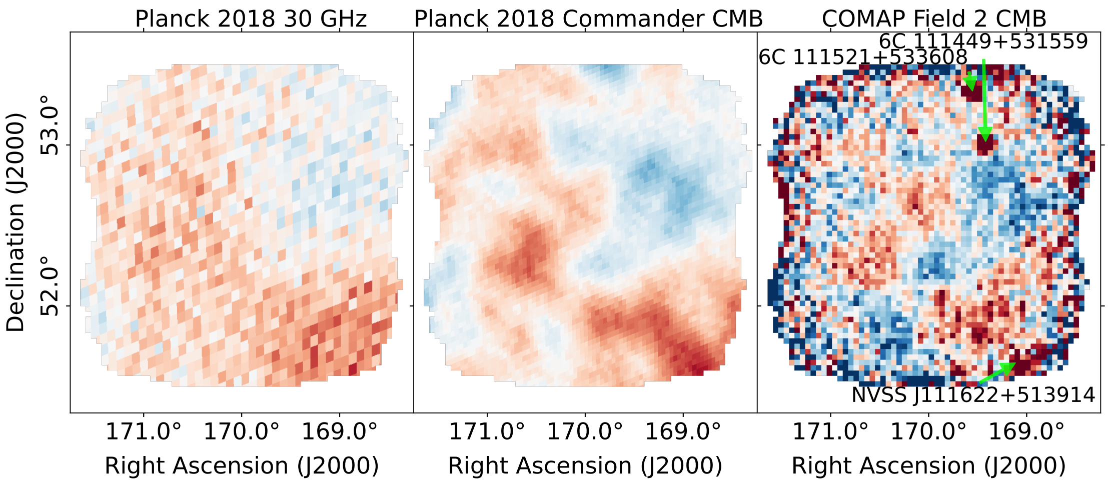


Figure 49: From left to right: Planck LFI 30 GHz map (Planck Collaboration II., 2020) reprojected and masked to the COMAP Field 2 footprint, the Planck 2018 Commander CMB map (Planck Collaboration IV., 2020), as well as the COMAP continuum leakage map. Three clear detections of point sources, as indicated by annotations. See Figs. 50, 51, and 52 as well as Table 4 for more point source detections.

(Planck Collaboration II., 2020; Planck Collaboration IV., 2020). The latter two are reprojected and masked to the same footprint as COMAP. As can be seen, the COMAP CMB estimate has a much higher resolution than the Planck LFI 30 GHz map, which is dominated by the Planck LFI 30 arcmin beam. Compared to the Planck 2018 Commander CMB temperature map, where most of the sensitivity comes from the 5 arcmin beam of the Planck HFI instrument, we clearly see that COMAP and Planck see the same sky signal. We can also see several strong point source detections in the COMAP map, as indicated by arrows in Fig. 49.

To show the high level of agreement between the COMAP and Planck CMB estimates in all three fields, we plot the COMAP continuum leakage maps with contours of the Planck Commander CMB map for all three fields in Figs. 50–52. Because the units of the COMAP continuum leakage are somewhat arbitrary because of the way the frequency channels are averaged, we perform a rough visual calibration (with a simple scaling and offset parameter) to the Planck Commander CMB map. In all three fields, there is a strong correlation between the calibrated COMAP and Planck Commander CMB maps. Additionally, we include contours of the NRAO VLA Sky Survey (NVSS) 1.4 GHz continuum survey (Condon et al., 2002) to cross-match the COMAP point sources to an external radio dataset.

As mentioned, we see several point sources detected in both COMAP and NVSS, as shown in Figs. 50–52. To identify the point sources seen in our COMAP continuum leakage maps, we can perform a catalog search in the Condon et al. (2002) NVSS catalog. The identified NVSS sources, including both tentative and significant detections, are marked as numbers in Figs. 50–52 next to the respective NVSS contours. In Table 4, we provide an overview of these, including the NVSS identifier, celestial coordinates, and the 1.4 GHz spectral flux density measurement by NVSS.

We also include approximate estimates of the spectral flux density of the point sources as measured by COMAP in Table 4. These are found by maximizing the S/N of isolated point sources given the beam and noise properties of the telescope. This is given by convolving the map with the beam B and then the inverse of the noise covariance N^{-1}. From the resulting matched-filtered map, we can read off the spectral flux density per pixel in the map $\boldsymbol{m}$ as

$$\boldsymbol{S}_{30\,\mathrm{GHz}}^{\mathrm{COMAP}} = \frac{\mathsf{B}^T\mathsf{N}^{-1}\boldsymbol{m}}{\mathrm{diag}(\mathsf{B}^T\mathsf{N}^{-1}\mathsf{B})} \tag{3.58}$$

with corresponding uncertainties

$$\boldsymbol{\sigma}_{S_{30\,\mathrm{GHz}}^{\mathrm{COMAP}}} = \frac{1}{\mathrm{diag}(\mathsf{B}^T\mathsf{N}^{-1}\mathsf{B})}. \tag{3.59}$$

Taking the ratio between these two, we see that we have several firm detections with up to 10.7σ significance, as well as some more tentative sources measured below 2σ significance.[6]

[6] We perform the matched filter and spectral flux density estimation with the `pixell` package; https://github.com/simonsobs/pixell.

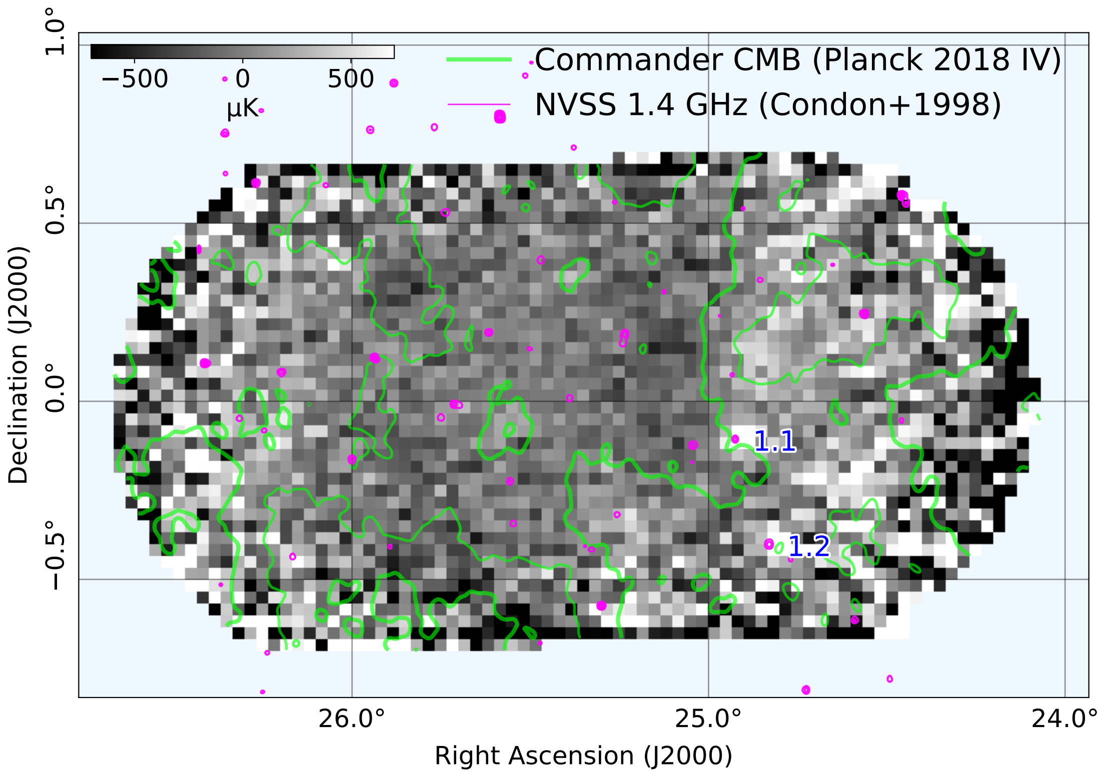


Figure 50: Continuum leakage map of COMAP Field 1, corresponding to the data volume of COMAP S2 (see Paper III). The Planck Commander CMB map (Planck Collaboration IV., 2020) and NVSS 1.4 GHz survey are, respectively, indicated as green and magenta contours. Point sources with a cross-match are marked with identifying numbers. In Table 4, we provide the Condon et al. (2002) NVSS identifier, equatorial coordinates, as well as 1.4 GHz NVSS and 30 GHz COMAP spectral flux density measurement of the identified sources. The COMAP leakage map is manually calibrated to the Planck CMB map.

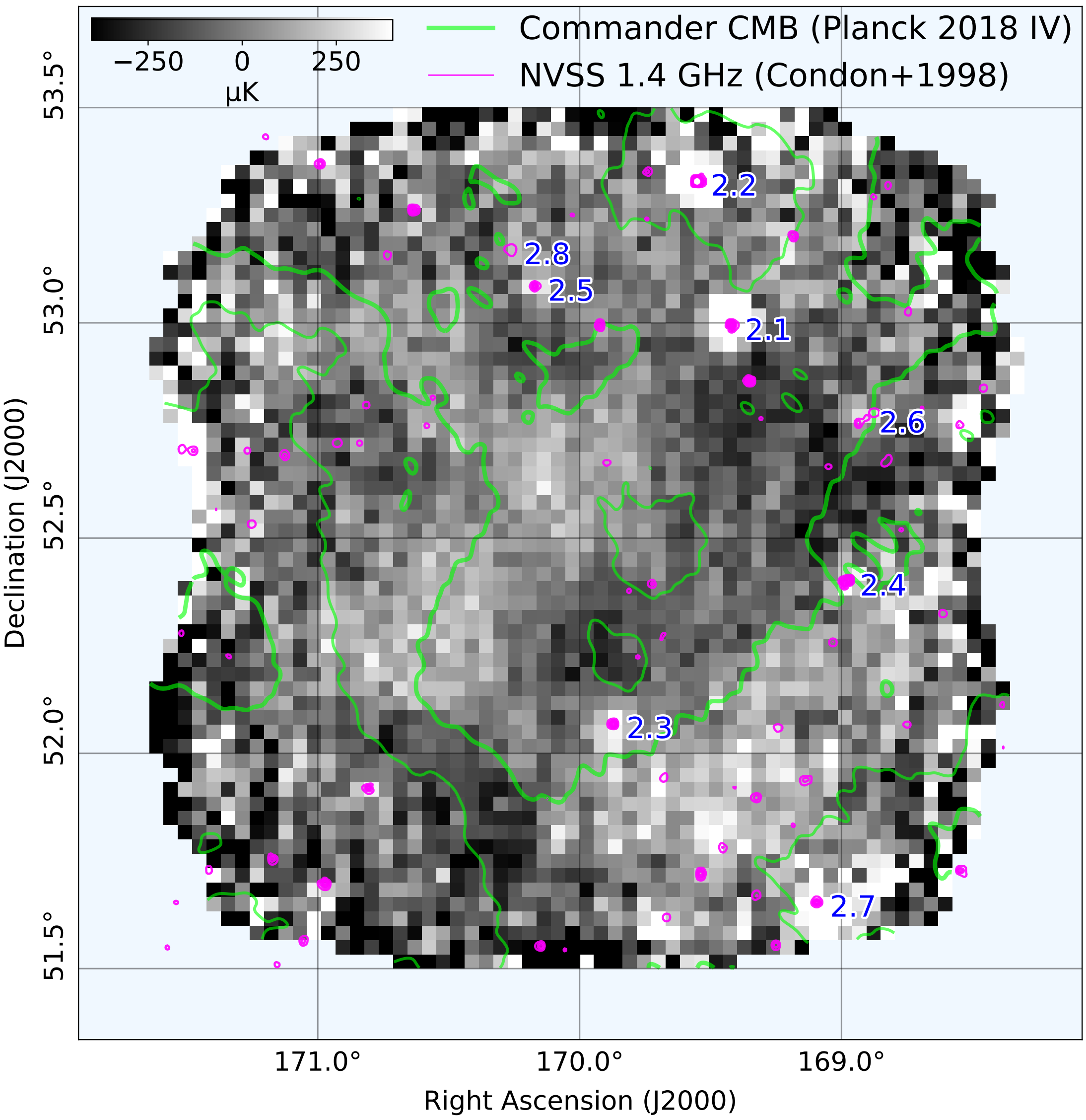


Figure 51: Same as Fig. 50 but for COMAP Field 2.

Although the CMB and point source detection in our COMAP Field 1–3 maps are not of the highest scientific value, they nevertheless are an interesting discovery and demonstrate the sensitivity of the experiment. It, however, also poses an interesting potential secondary science goal of the COMAP survey, namely, using the COMAP instrument for cosmological continuum science applications. As demonstrated by our results, the COMAP instrument is a perfect 30 GHz match to the Planck HFI instrument at higher frequencies because of the similar angular resolutions. Furthermore, having a high spectral resolution thus makes COMAP ideally suited to measure the spectra of, e.g., high-redshift Sunyaev-Zeldovich (SZ) clusters at 30 GHz.

The residual continuum is found in our COMAP S2 maps after processing the raw data through our low-level pipeline, as explained in Sec. 3.4. The individual frequency channels in the map of each field are then coadded using inverse variance weighting. The exact reason why we see a continuum residual in the frequency-

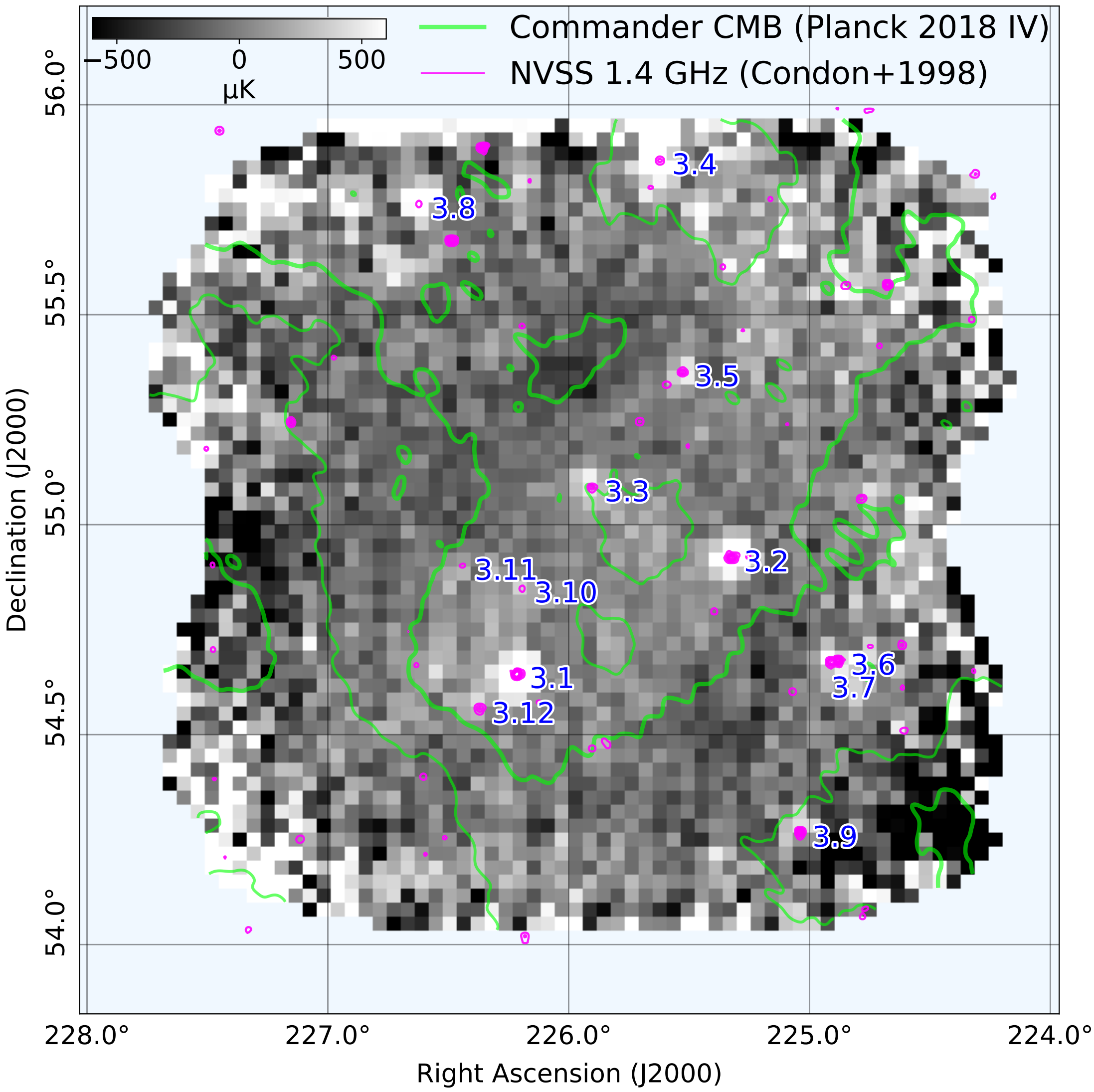


Figure 52: Same as Fig. 50 but for COMAP Field 3.

averaged COMAP maps is somewhat beyond the scope of this thesis and will be explained in detail in future COMAP publications and the upcoming Ph.D. thesis of Jonas Lunde. In simple terms, the continuum radiation can leak into a noise-weighted average of all channels due to the polynomial frequency filter (see Sec. 3.4.2) only fitting a low-ordered polynomial across frequencies in an individual sideband. Meanwhile, the sensitivity structure across the COMAP bandpass has several system temperature spikes (see Fig. 18). The result is that the CMB and point source continuum survive as a weak leakage effect if the data are subsequently weighted and averaged by this system temperature profile.

COMAP ID	NVSS ID	Right Ascension [J2000]	Declination [J2000]	$S^{\mathrm{NVSS}}_{1.4\,\mathrm{GHz}}$ [mJy]	$\sigma_{S^{\mathrm{NVSS}}_{1.4\,\mathrm{GHz}}}$ [mJy]	$S^{\mathrm{COMAP}}_{30\,\mathrm{GHz}}$ [mJy]	$\sigma_{S^{\mathrm{COMAP}}_{30\,\mathrm{GHz}}}$ [mJy]	COMAP S/N
Field 1								
1.1	013942-000619	24.925°	−0.105°	43.8	1.4	54.6	13.9	3.9
1.2	013919-002403	24.830°	−0.401°	68.8	2.1	50.0	25.3	2.0
Field 2								
2.1	111740+525937	169.418°	52.994°	209.1	6.3	90.3	9.3	9.7
2.2	111811+531944	169.549°	53.329°	919.2	32.5	96.7	14.3	6.8
2.3	111929+520407	169.871°	52.069°	139.4	4.2	25.7	8.0	3.2
2.4	111555+522400	168.979°	52.400°	325.0	10.2	24.9	12.0	2.1
2.5	112041+530505	170.171°	53.085°	101.9	3.1	17.5	9.3	1.9
2.6	111537+524643	168.906°	52.779°	15.2	0.6	24.0	12.8	1.9
2.7	111622+513914	169.094°	51.654°	125.9	4.4	33.9	23.0	1.5
2.8	112103+531011	170.263°	53.170°	84.0	3.5	9.4	10.2	0.9
Field 3								
3.1	150451+543839	226.213°	54.644°	376.2	11.3	91.8	8.6	10.7
3.2	150117+545519	225.323°	54.922°	196.3	7.0	71.6	9.6	7.5
3.3	150336+550519	225.901°	55.089°	62.8	1.9	32.0	7.6	4.2
3.4	150229+555204	225.621°	55.868°	34.8	1.1	62.4	17.6	3.5
3.5	150206+552146	225.527°	55.363°	89.4	2.7	27.4	9.2	3.0
3.6	145931+544033	224.880°	54.676°	256.7	7.7	38.8	14.0	2.8
3.7	145938+544018	224.909°	54.672°	108.3	3.9	36.9	13.6	2.7
3.8	150629+554546	226.622°	55.763°	21.8	1.1	44.7	17.6	2.5
3.9	150009+541559	225.038°	54.267°	173.6	6.2	33.9	18.7	1.8
3.10	150446+545053	226.193°	54.848°	19.4	0.7	11.5	7.9	1.5
3.11	150545+545407	226.441°	54.902°	22.8	1.5	8.5	9.0	0.9
3.12	150528+543340	226.369°	54.561°	79.0	2.4	7.3	9.7	0.7

Table 4: Detection of point sources in the COMAP continuum residual maps and their corresponding NVSS 1.4 GHz catalog cross-match. Tentative detections with S/N < 2 are marked with gray shading. For each field, 1–3, the point sources are indicated as numbers in the maps seen in Fig. 50, 51 and 52. The NVSS identifier, coordinates, and spectral flux density with uncertainties, S_ν and σ_{S_ν}, are gathered from the NVSS 1.4 GHz catalog of Condon et al. (2002). The point sources are sorted according to the COMAP spectral flux density estimate's signal-to-noise ratio $S/N = S^{\mathrm{COMAP}}_{30\,\mathrm{GHz}}/\sigma_{S^{\mathrm{COMAP}}_{30\,\mathrm{GHz}}}$.

Chapter 4

Preparing for the Future: From Linear to Global Iterative Data Analysis

Most of the work in this thesis has been done on data analysis for the COMAP experiment. However, in Papers VI and VII, we also explore some algorithmic improvements developed for CMB analysis that can, in principle, be adapted to a LIM experiment such as COMAP. As described in Sec. 3.4, the COMAP pipeline is an example of a classical linear pipeline in which the data are filtered, calibrated, and binned into maps before inference is performed using power spectra. These pipelines are often a good first approach to data analysis while still learning about the properties of the data. However, there are more well-motivated end-to-end pipeline architectures, such as those built on Bayesian techniques like Gibbs sampling (Geman & Geman, 1984; Wandelt et al., 2004; Eriksen et al., 2004; O'Dwyer et al., 2004). One of these pipelines is called `Commander3` and was originally developed for the BeyondPlanck collaboration (BeyondPlanck Collaboration et al., 2023) and is currently further generalized to leverage all cosmological datasets jointly in the Cosmoglobe project (see, e.g, Watts et al., 2023a). The `Commander3` framework is built on earlier iterations of the `Commander` code, such as `Commander1` and `Commander2`. `Commander1` was originally used in the Planck analysis (Eriksen et al., 2004, 2008; Seljebotn et al., 2019) and represents the starting point of the Cosmoglobe project (Planck Collaboration X., 2016; Planck Collaboration IV., 2020). These Bayesian end-to-end frameworks can explore the system's full posterior probability space of parameters and even reveal all non-trivial correlations between parameters. In contrast, a simple linear pipeline (such as that of COMAP) often treats parameters in the time and map domains separately, and knowledge of the correlation between these is lost. This then results in underestimated parameter uncertainties. The `Commander3` code is currently the only full global end-to-end analysis framework that can map out the posterior probability from the time-domain all the way to cosmic parameters.

We begin this chapter by giving an overview of `Commander3`. Thereafter, we present the work done for Paper VI, which used simulations to validate the mapmaking, gain, and correlated noise estimation in the `Commander3`. Lastly, we consider our contribution to Paper VII in which we use maximum likelihood mapmaking to account for the bolometer transfer function of the Planck HFI

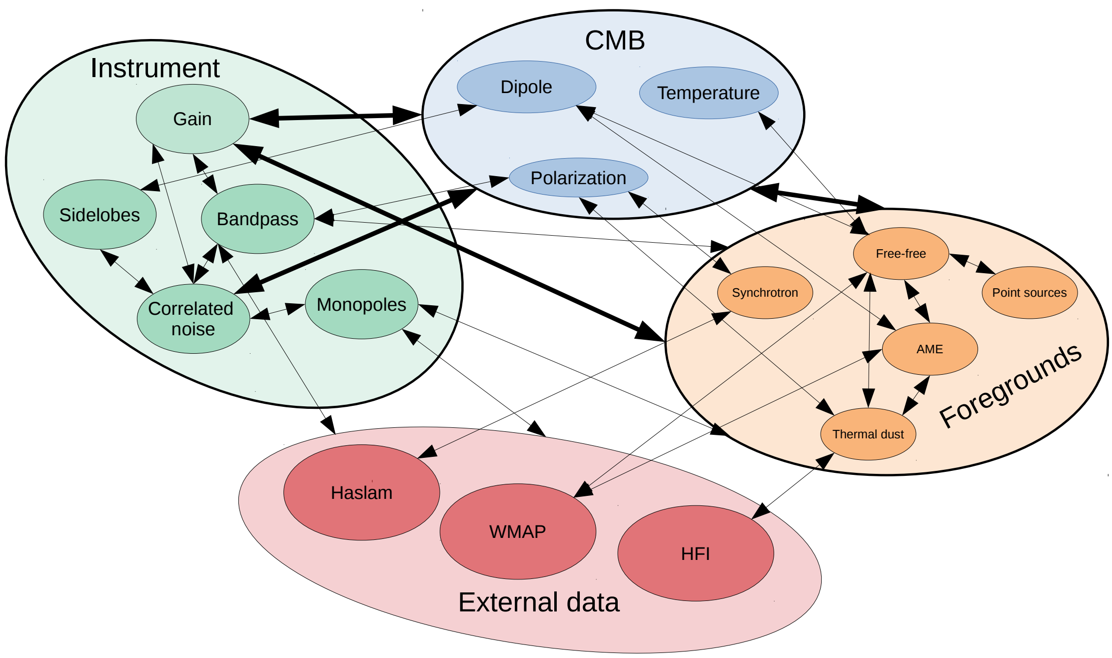


Figure 53: Overview of central parameter and some external datasets used in the BeyondPlanck analysis. Arrows indicate some known inter-dependencies between parameters and datasets, illustrating the complex degeneracies of the system. The figure is taken from the BeyondPlanck Collaboration et al. (2023).

bolometers. Specifically, we present work on a simplified and independent validation of the main algorithm in the paper. In the future, the work in these two papers can be adapted to the COMAP experiment. It will become important for COMAP to become a high-precision cosmology survey once the first cosmic CO detection is in sight.

4.1 Global Bayesian data analysis

The first step in an iterative Bayesian sampler like `Commander3` is to write down a parametric data model that captures all aspects of the real data. In Paper VI, we consider Gibbs sampling applied to the Planck LFI data and can be written as

$$\begin{aligned} d_{j,t} =& g_{j,t}\mathsf{P}_{tp,j}\left[\mathsf{B}^{\mathrm{symm}}_{pp',j}\sum_{c}\mathsf{M}_{cj}(\beta'_p,\Delta^j_{\mathrm{bp}})a^c_{p'} + \mathsf{B}^{4\pi}_{i,t}s^{\mathrm{orb}}_j + \mathsf{B}^{\mathrm{asymm}}_{j,t}s^{\mathrm{fsl}}_t\right] \\ &+ a_{1\,\mathrm{Hz}}s^{1\,\mathrm{Hz}}_j + n^{\mathrm{corr}}_{j,t} + n^{\mathrm{w}}_{j,t}, \end{aligned} \tag{4.1}$$

where indices p, c, t, and j indicate a pixel on the sky, an astrophysical component (like the CMB, free-free, synchrotron radiation, etc.), time step, and radiometer, respectively. Here $d_{j,t}$ represents the TOD, $g_{j,t}$ is the instrumental gain, $\mathsf{P}_{tp,j}$ denotes the pointing matrix. The beam convolution is denoted as $\mathsf{B}_{pp',j}$, for the symmetric main beam and asymmetric far sidelobes of the full 4π beam response. The mixing matrix $\mathsf{M}_{cj}(\beta_p, \Delta_{\mathrm{pb}})$ describes the spectral response of an astrophysical component c as observed in radiometer j given a bandpass correction

Δ_{bp} and spectral indices β_p. The corresponding amplitude of component c is given by a_p^c in pixel p. Furthermore, the orbital CMB dipole is given as $s_{j,t}^{\mathrm{orb}}$, the far sidelobes signal is denoted by $s_{j,t}^{\mathrm{fsl}}$ and $s_{j,t}^{1\,\mathrm{Hz}}$ corresponds to the contributions from electronic 1 Hz spikes with amplitude $a_{1\,\mathrm{Hz}}$. The correlated and uncorrelated (white) instrumental noise is denoted, respectively, by $n_{j,t}^{\mathrm{corr}}$ and $n_{j,t}^{\mathrm{w}}$.

The sum in Eq. (4.1) sums over all the astrophysical components, i.e., the CMB, synchrotron, free-free, AME, thermal dust, and point sources, to obtain a full sky model (see Andersen et al., 2023; Svalheim et al., 2023). Meanwhile, the noise properties of the data $d_{j,t}$ are characterized by the noise covariance matrix $\mathsf{N}^{\mathrm{corr}} = \langle \boldsymbol{n}^{\mathrm{corr}} (\boldsymbol{n}^{\mathrm{corr}})^T \rangle$. It has a simple diagonal form in Fourier space given by the power spectral density (PSD)

$$P(f) = \sigma_0 \left[1 + \left(\frac{f}{f_{\mathrm{knee}}} \right)^{\alpha} \right] + A_{\mathrm{p}} \exp \left[-\frac{1}{2} \left(\frac{\log_{10} f - \log_{10} f_{\mathrm{p}}}{\sigma_{\mathrm{dex}}} \right)^2 \right]. \tag{4.2}$$

This noise PSD, described in detail by Ihle et al. (2023), combines a $1/f$ spectrum (that we have already seen earlier in the COMAP data model; Eq. 3.7) with a log-normal component with fixed width σ_{dex} and peak f_{p}. The free parameters of the noise model in Eq. (4.2) can be collected into $\xi_n = \{\sigma_0, \alpha, f_{\mathrm{knee}}, A_{\mathrm{p}}\}$, and can later be sampled at the same time as a combined parameter vector.

The next step in the BeyondPlanck analysis is to map out the posterior probability space of the free parameters in Eq. (4.1). The posterior probability distribution of all free parameters, collected into $\boldsymbol{\omega}$, is given by Bayes' theorem

$$P(\boldsymbol{\omega}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{\omega}) P(\boldsymbol{\omega})}{P(\boldsymbol{d})} \propto \mathcal{L}(\boldsymbol{\omega}) P(\boldsymbol{\omega}). \tag{4.3}$$

The term $P(\boldsymbol{d}|\boldsymbol{\omega}) = \mathcal{L}(\boldsymbol{\omega})$ is the likelihood and refers to the probability of the data given the parameters, $P(\boldsymbol{\omega})$ denotes the prior and quantifies the prior information we have on the parameters from theory or other experiments (see BeyondPlanck Collaboration et al., 2023, for details on priors used `Commander3`). The evidence term $P(\boldsymbol{d})$ is normally omitted as it is simply a normalization factor when considering parameter inference.

Assuming Gaussian noise properties, we can write the likelihood in the following form:

$$-2 \ln \mathcal{L}(\boldsymbol{\omega}) = (\boldsymbol{d} - \boldsymbol{s}^{\mathrm{tot}}(\boldsymbol{\omega}))^T \mathsf{N}_{\mathrm{w}}^{-1} (\boldsymbol{d} - \boldsymbol{s}^{\mathrm{tot}}(\boldsymbol{\omega})), \tag{4.4}$$

where $\boldsymbol{s}^{\mathrm{tot}}(\boldsymbol{\omega})$ constitutes all contributions to Eq. (4.1) but the white noise $\boldsymbol{n}_{\mathrm{w}}$, and N_{w} is the white noise covariance matrix.

As we can see from Eqs. (4.4) and (4.1), the model and corresponding likelihood function are dependent on a plethora of parameters that could be correlated in highly non-trivial ways. In Fig. 53, we see a diagram illustrating a few known correlations and degeneracies. However, mapping the posterior probability space using some Markov Chain Monte Carlo (MCMC) algorithm is not trivial simply because the number of parameters is so large, with up to millions of free parameters (Wandelt et al., 2004; Eriksen et al., 2004; O'Dwyer et al., 2004). Currently, the only viable MCMC method for a problem of these dimensions is Gibbs sampling (Geman & Geman, 1984; Wandelt et al., 2004; Eriksen et al., 2004), which is

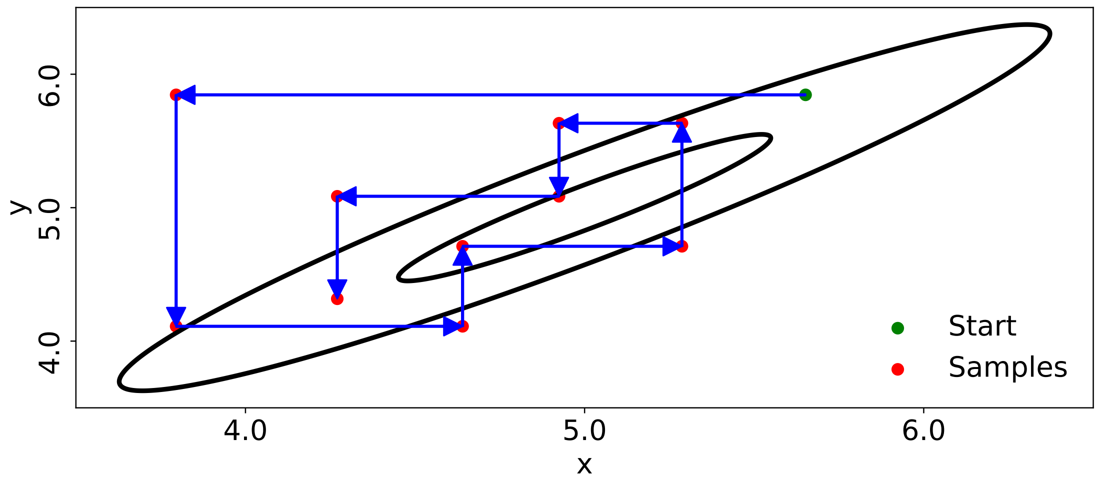


Figure 54: Two-dimensional example of a Gibbs chain sampling a two-variate Gaussian distribution (indicated by the 1σ and 2σ contours). The sampler starts at the green point and moves along conditional probabilities as indicated by the arrows between subsequent samples (red).

the main algorithm `Commander3` is built around (see BeyondPlanck Collaboration et al., 2023, for details).

Gibbs sampling is an MCMC algorithm where the free parameters are broken down into groups that can be sampled jointly in a single step. We then sample the parameters from the conditional distribution for each group, keeping all other parameter groups fixed. In this way, we can sample thousands of parameters, e.g., all gain parameters $\boldsymbol{g}$, as a vector instead of each element of the vector individually. This is what makes Gibbs sampling so efficient for high-dimensional probability distributions, as we consider here.

Algorithmically, we can write the Gibbs sampler as

$$\boldsymbol{g} \leftarrow P(\boldsymbol{g} \,|\boldsymbol{d}, \xi_n, \boldsymbol{a}_{1\,\mathrm{Hz}}, \Delta_{\mathrm{bp}}, \boldsymbol{a}, \beta, C_\ell), \tag{4.5}$$

$$\boldsymbol{n}_{\mathrm{corr}} \leftarrow P(\boldsymbol{n}_{\mathrm{corr}}|\boldsymbol{d}, \boldsymbol{g}, \xi_n, \boldsymbol{a}_{1\,\mathrm{Hz}}, \Delta_{\mathrm{bp}}, \boldsymbol{a}, \beta, C_\ell), \tag{4.6}$$

$$\xi_n \leftarrow P(\xi_n \,|\boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\mathrm{corr}}, \boldsymbol{a}_{1\,\mathrm{Hz}}, \Delta_{\mathrm{bp}}, \boldsymbol{a}, \beta, C_\ell), \tag{4.7}$$

$$\boldsymbol{a}_{1\,\mathrm{Hz}} \leftarrow P(\boldsymbol{a}_{1\,\mathrm{Hz}}|\boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\mathrm{corr}}, \xi_n, \Delta_{\mathrm{bp}}, \boldsymbol{a}, \beta, C_\ell), \tag{4.8}$$

$$\Delta_{\mathrm{bp}} \leftarrow P(\Delta_{\mathrm{bp}} \,|\boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\mathrm{corr}}, \xi_n, \boldsymbol{a}_{1\,\mathrm{Hz}}, \boldsymbol{a}, \beta, C_\ell), \tag{4.9}$$

$$\beta \leftarrow P(\beta \,|\boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\mathrm{corr}}, \xi_n, \boldsymbol{a}_{1\,\mathrm{Hz}}, \Delta_{\mathrm{bp}}, C_\ell), \tag{4.10}$$

$$\boldsymbol{a} \leftarrow P(\boldsymbol{a} \,|\boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\mathrm{corr}}, \xi_n, \boldsymbol{a}_{1\,\mathrm{Hz}}, \Delta_{\mathrm{bp}}, \beta, C_\ell), \tag{4.11}$$

$$C_\ell \leftarrow P(C_\ell \,|\boldsymbol{d}, \boldsymbol{g}, \boldsymbol{n}_{\mathrm{corr}}, \xi_n, \boldsymbol{a}_{1\,\mathrm{Hz}}, \Delta_{\mathrm{bp}}, \boldsymbol{a}, \beta), \tag{4.12}$$

where each line represents a sampling step, denoted by $\leftarrow$, from the conditional probability distribution of each parameter (group) of the model (BeyondPlanck Collaboration et al., 2023). Intuitively, we can think of each step in Eqs. (4.5)–(4.12) as moving along individual coordinate axes at a time while keeping the rest of the parameters fixed. This is illustrated in Fig. 54, where we show a low-dimensional sketch of a Gibbs chain.

Furthermore, we do not have to stop at a single experiment. The true power of a global Bayesian Gibbs sampler like `Commander3` is unfolded when combining multiple experiments, each covering different frequencies and resolutions in the sky, all with different systematic sources of error. In fact, this is what the Cosmoglobe project is aiming to achieve (see Watts et al., 2023a,b; Eskilt et al., 2023; Watts et al., 2024b,c,a; San et al., 2024, for the current state of the project). In this way, different experiments can complement each other, and ever-deeper systematic errors can be rooted out. The power of Gibbs sampling is, therefore, unmatched by any classical pipeline like the current implementation of the COMAP pipeline. Adapting the `Commander` framework to COMAP will be an important future step toward becoming a high-precision cosmology survey, and other LIM and CMB surveys will all mutually benefit from a joint analysis. Currently, the main challenge of incorporating an experiment like COMAP into this framework is the enormous data volume of the spectrographic time-ordered data, and we thus likely also need some additional large-scale parallelization improvements for this to work. This is the goal of the `Commander4` code, of which the first versions are currently being developed under a recent European Research Council Advanced Grant. In this framework, it will be possible to jointly analyze large CMB datasets like those of Simon's Observatory and LiteBIRD, but also possibly LIM data such as COMAP and other experiments mentioned in Ch. 2.

4.2 Simulations and validation

In this section, we will consider pipeline validation of three central parts of the `Commander3` code, namely gain and correlated noise estimation and mapmaking. The validation of `Commander3` is one of the two topics covered in Paper VI, and the one we worked more extensively on for this thesis. The other part of Paper VI explores the difference between Bayesian posterior and frequentist prior simulations and their uses. Therefore, we will only consider the pipeline validation part of Paper VI, and refer the interested reader to the paper for more details on the data simulation.

4.2.1 Pipeline validation

The goal of the pipeline validation presented in Paper VI is to validate central algorithms in `Commander3`, namely the gain and noise estimation and mapmaking. These algorithms are described in detail by Gjerløw et al. (2023), Ihle et al. (2023), and Keihänen et al. (2023), respectively, and we refer the interested reader to their work as the algorithms themselves are beyond the scope of this thesis.

To perform a validation of these algorithms, we use a set of simulated Planck LFI data, limiting ourselves to only the LFI 30 GHz channel and consider a relatively small data volume of around 10,000 PIDs corresponding to about a year of data (Planck Collaboration I., 2014). Due to the simplified dataset, we can produce Gibbs chains of up to 10,000 samples in a matter of a few days to weeks instead of waiting several months, as was the case for the full BeyondPlanck re-analysis of Planck LFI (Galloway et al., 2023).

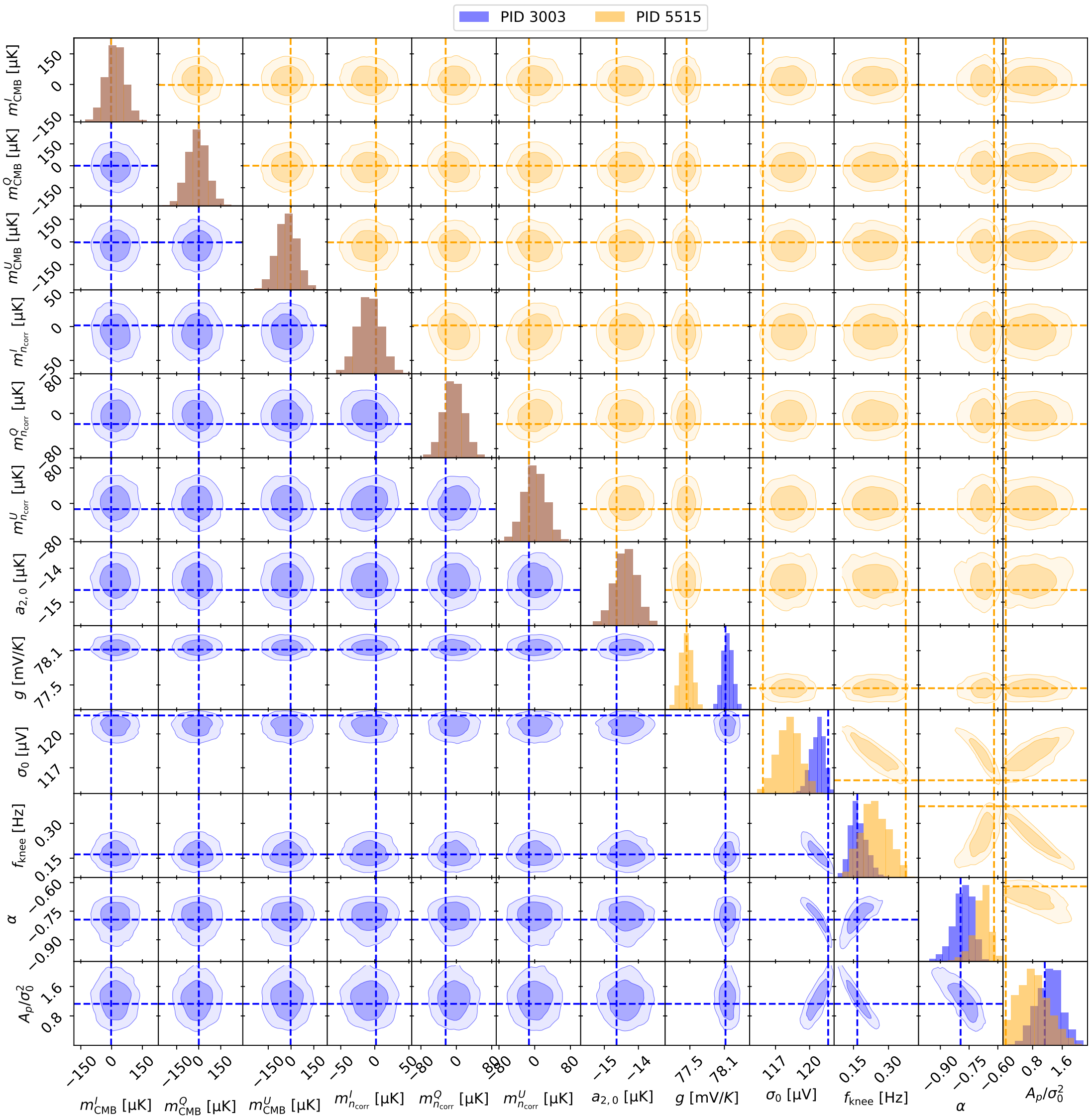


Figure 55: Estimated posterior distribution of several parameters for two different PIDs, respectively indicated by blue and yellow 68 % and 95 % confidence intervals. The two PIDs (blue and yellow) represent scans where the parameters are recovered within 1σ while the other represents a more suboptimal recovery within 2σ. Dashed lines indicate the true input parameters used to generate the simulated data. The parameters, from left to right along the horizontal axis correspond to (1)–(3) Stokes I, Q and U of a random pixel of the CMB temperature map m_{CMB}; (4)–(6) Stokes I, Q and U of a the correlated noise map pixel m_{corr}; (7) the CMB quadrupole amplitude $a_{2,0}$; (8) the gain g; (9)–(12) the noise parameters $\sigma_0, \alpha, f_{\mathrm{knee}}, A_{\mathrm{p}}$. The figure is taken from Paper VI.

Because we are only interested in validating low-level algorithms, we consider simplification of the full data model in Eq. (4.1) of the form:

$$d^{\mathrm{sim}}_{i,t} = g_{j,t}\mathsf{P}_{tp,j}\mathsf{B}^{\mathrm{symm}}_{pp',j}a^{\mathrm{cmb}}_{p'} + \mathsf{B}^{\mathrm{asymm}}_{pp',j}s^{\mathrm{orb}}_{j,t} + n^{\mathrm{corr}}_{j,t} + n^{\mathrm{w}}_{j,t} \quad = s^{\mathrm{tot}}_{j,t} + n^{\mathrm{corr}}_{j,t} + n^{\mathrm{w}}_{j,t}, \quad (4.13)$$

only including the CMB while ignoring other astrophysical effects, far sidelobes etc. As such, this does not validate the component separation, as only one sky

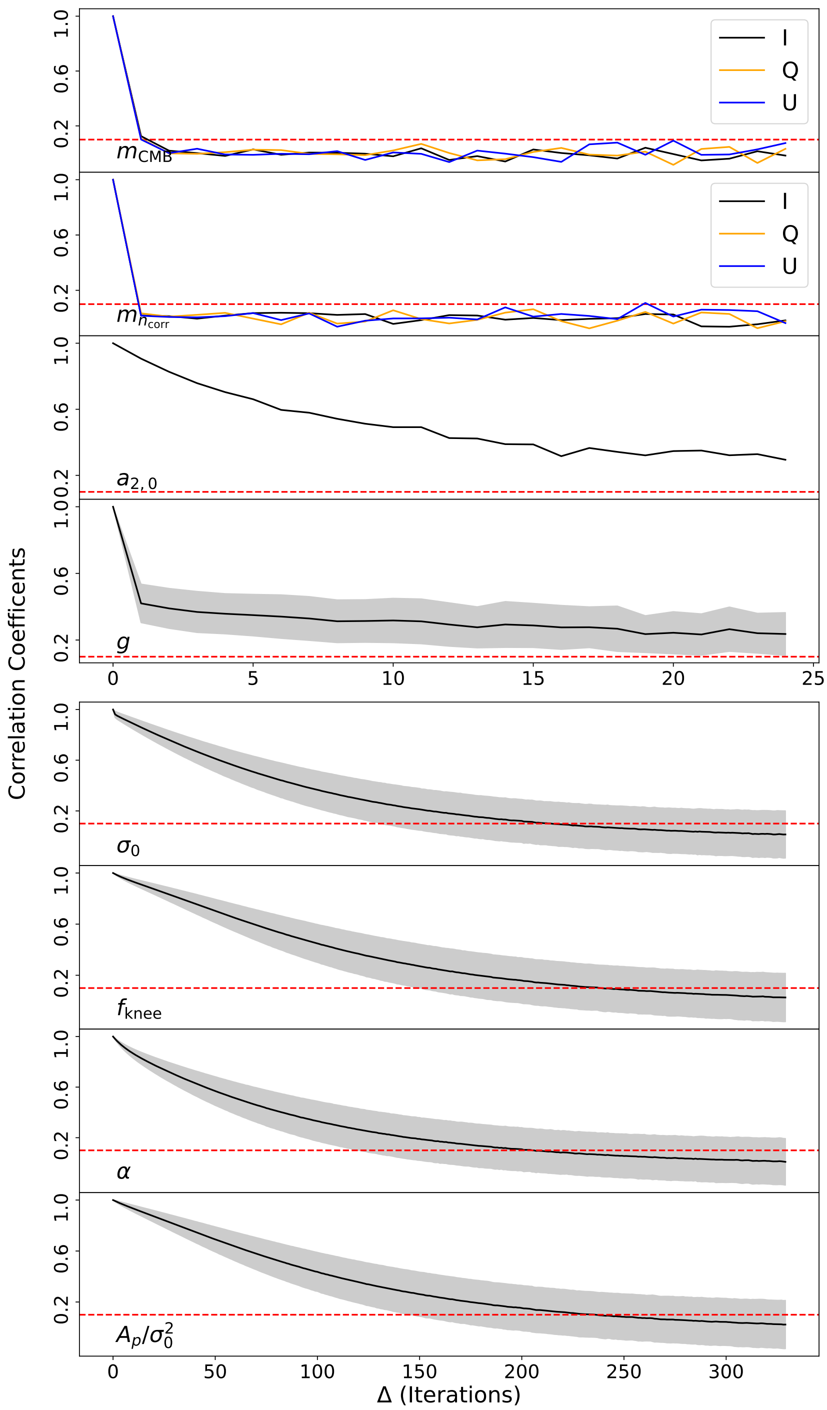


Figure 56: Auto-correlation function $\rho(\omega)$ (see Eq. (4.14)) as a function of chain separation Δ for several chosen parameters of the data model in Eq. (4.13) computed from a 10,000 sample Gibbs chain. From top to bottom, the panels correspond to some pixel value of the CMB component map $m_{\rm CMB}$ for both Stokes parameters I, Q and U; one pixel of the correlated noise map $m_{n_{\rm corr}}$ in all three Stokes parameters; the quadrupole moment $a_{2,0}$ of the CMB temperature; the average gain g across PIDs; as well as the noise PSD parameters σ_0, $f_{\rm knee}$, α and $A_{\rm p}/\sigma_0^2$. In the bottom five panels, solid black lines represent the PID averaged values, while gray bands represent the 1σ spread across PIDs. The dashed red line represents the 0.1 auto-correlation commonly used to define the correlation length of the parameter. The figure is adapted from Paper VI.

component and frequency channel are included. The realizations of the CMB map, a_p^{cmb}, are generated using the HEALPix (Górski et al., 2005) `synfast` tool. Specifically, they are drawn from a best-fit Planck 2018 ΛCDM model (Planck Collaboration V., 2020). For the simulations, we use known instrumental input parameters drawn from the BeyondPlanck ensemble of BeyondPlanck Collaboration et al. (2023), and the noise is drawn from the noise model we presented in Eq. (4.2) independently for each PID.

Using these simulations, a Gibbs chain of 10,000 samples is produced using the `Commander3` code described in Sec. 4.1. The main goal is to find out how well the input parameters are reconstructed and to map out uncertainties and correlations of parameters. In Fig. 55, we see the marginalized posterior probability distributions mapped out by our Gibbs chain. Specifically, we show the posterior distribution for two PIDs. One in which most input parameters are recovered within the 68 % confidence interval (blue). At the same time, the other corresponds to a less well-behaved PID in which the parameters are recovered only within the 95 % confidence interval (yellow). The two PIDs should, therefore, cover most of the possible outcomes of the recovery procedure. We can see that all parameters are generally recovered well, even in the problematic PID. However, we note that the noise parameters σ_0, α, f_{knee}, $A_{\mathrm{p}}/\sigma_0^2$ show clear signs of mutual correlations as evidenced by their skewed contours.

To judge whether or not the sample chains have burned in properly, we can look at the auto-correlation function

$$\rho_\omega(\Delta) = \left\langle \left(\frac{\omega^i - \mu_\omega}{\sigma_\omega}\right)\left(\frac{\omega^{i+\Delta} - \mu_\omega}{\sigma_\omega}\right)\right\rangle, \tag{4.14}$$

for Gibbs sample number i and chain sample offset Δ. This will quantify the correlation between samples in the chain as a function of their separation.

In Fig. 56 we can see the auto-correlation function for a set of chosen parameters: a pixel of the Stokes I, Q and U CMB map m_{CMB}; the corresponding correlated noise map pixel m_{corr}; the quadrupole amplitude of the CMB temperature map; the gain g as well as the noise parameters σ_0, α, f_{knee} and $A_{\mathrm{p}}/\sigma_0^2$ of a single PID. We see that the CMB and correlated noise maps converge to auto-correlations below 0.1 within only a few samples, as these are dominated by only white noise in single pixels. However, the quadrupole amplitude $a_{2,0}$ and gain g need considerably longer because they are correlated to each other. Additionally, the gain relies on the CMB as a calibration source. This degeneracy is mitigated in the full BeyondPlanck analysis, as all Planck LFI channels, as well as WMAP data, are analyzed together (Gjerløw et al., 2023; Basyrov et al., 2023).

Meanwhile, the noise parameters have much longer correlation lengths. This is most likely caused by internal degeneracies within the noise PSD itself, as noted by Ihle et al. (2023) and the marginal posterior distributions shown in Fig. 55. However, as we find in Paper VI, the noise PSD is relatively insensitive to changes in the noise parameters, and the degeneracies do not cause any biases in any of the other more important parameters of the system.

In general, we can conclude that the `Commander3` framework seems to work as intended. Our validation performed in Paper VI shows that we can recover the input parameters used to simulate the data. Importantly, we can also map out the correlation between the parameters, which would otherwise remain difficult to estimate in a different framework. Thus, moving towards a similar approach in a future rendition of the COMAP pipeline will be an important goal and aid COMAP in becoming a high-precision experiment like modern CMB surveys.

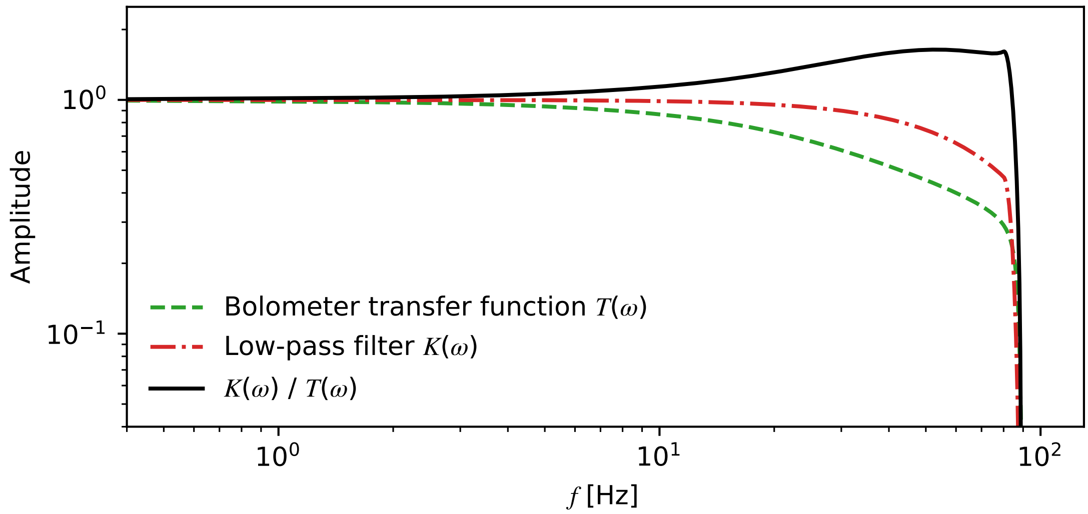


Figure 57: Bolometer transfer function of Planck HFI $T(\omega)$ (green), the Planck HFI low-pass filter $K(\omega)$ used to suppress noise on short time scales (red), as well as the ratio between the two (black). Note that the bolometer transfer function also has a complex component. Courtesy of Jonas Lunde, see Paper VII.

4.3 Optimal maximum likelihood mapmaking

The seventh and last paper (Paper VII) of this thesis is concerned with optimal mapmaking for experiments using bolometers. Although COMAP, the main subject of this thesis, does not use bolometers, some existing and upcoming LIM experiments will likely use bolometers. Thus, this work will be relevant for all bolometric (or detectors of similar properties) CMB, LIM, and other experiments. Specifically, our contribution to Paper VII is to use a simplified 1D toy model to build intuition for the main bolometer transfer function mapmaker proposed in the paper and help verify the observations made.

To summarize the work, we start with a data model for a bolometric detector like the one used by Planck HFI. Bolometers work by measuring the temperature-dependent resistance across a radiation absorber to infer the intensity of the light. Thus, a bolometric detector will heat up by the radiation of a source as it scans across it and registers a signal while it is cooling down (see Richards, 1994, for review on bolometers). As a result, bolometers have an associated time constant that will smooth the signal along the scanning direction. This can be seen in a simulated point source experiment from Paper VII in Fig. 58.

We can write an expression for the data on the form

$$\boldsymbol{d} = \mathsf{T}\boldsymbol{s} + \boldsymbol{n}, \tag{4.15}$$

where the bolometer transfer function $\mathsf{T} = \mathsf{F}^{-1}\mathsf{T}(\omega)\mathsf{F}$, with Fourier transform F, represents the smoothing of the bolometer time constant. The noise is given by $\boldsymbol{n}$. The bolometer transfer function $\mathsf{T}(\omega)$ can be seen in Fig. 57. In the original Planck HFI analysis (see Planck Collaboration VII, 2014, 2016), the bolometer transfer function was simply deconvolved in the time domain so that

$$\mathsf{T}^{-1}\boldsymbol{d} = \mathsf{T}^{-1}(\mathsf{T}\boldsymbol{s} + \boldsymbol{n}) = \boldsymbol{s} + \mathsf{T}^{-1}\boldsymbol{n}. \tag{4.16}$$

can be used to obtain an unbiased map of the signal $\boldsymbol{s}$. The problem with this deconvolution is that $T(\omega)$ goes towards zero at short time scales, and hence the term $\mathsf{T}^{-1}\boldsymbol{n}$ will increase the noise in the system. To deal with this, the Planck HFI analysis additionally applied a low-pass filter K to modulate the noise. This low-pass filter function is shown in Fig. 57. Then, the time stream

$$\mathsf{K}\mathsf{T}^{-1}\boldsymbol{d} = \mathsf{K}\mathsf{T}^{-1}(\mathsf{T}\boldsymbol{s} + \boldsymbol{n}) = \mathsf{K}\boldsymbol{s} + \mathsf{K}\mathsf{T}^{-1}\boldsymbol{n}. \tag{4.17}$$

is binned up into maps using the mapmaking equation (Tegmark, 1997) like we have seen previously in Sec. 3.4.3 on the COMAP mapmaker,

$$\mathsf{P}^T\mathsf{N}^{-1}\mathsf{P}\hat{\boldsymbol{m}}_{\mathrm{trad}} = \mathsf{P}^T\mathsf{N}^{-1}\boldsymbol{d}. \tag{4.18}$$

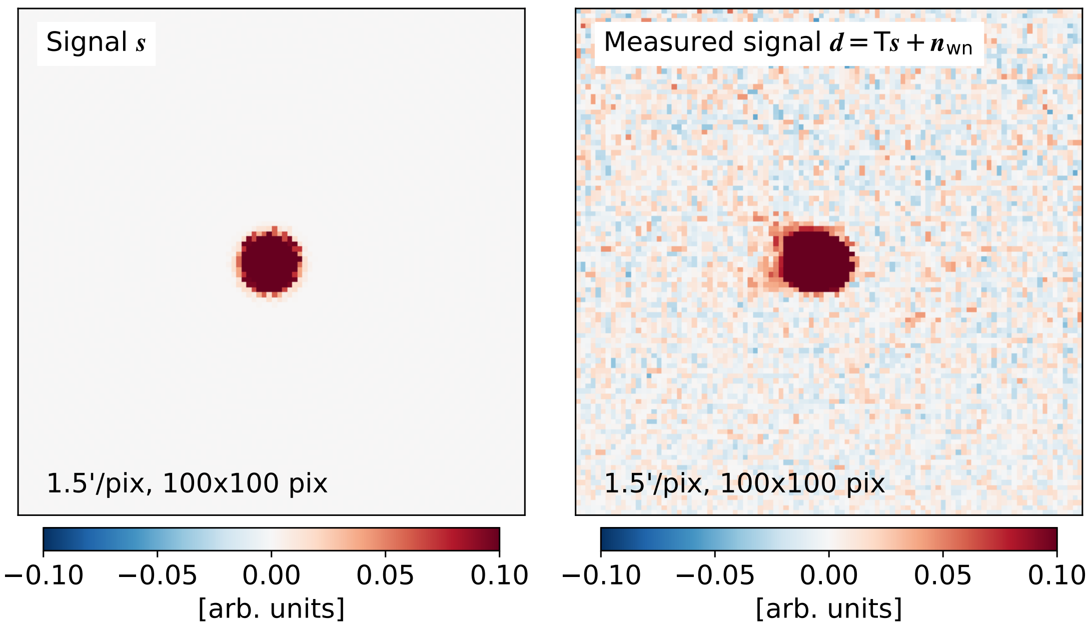


Figure 58: Simulated Gaussian point source $\boldsymbol{s}$ (*left*) and how the point source is measured by the Planck HFI bolometers $\boldsymbol{d} = \mathsf{T}\boldsymbol{s} + \boldsymbol{n}$ (*right*). We can see how the observed point source is smeared along the scanning direction by the finite time constant of the bolometer transfer function T. Courtesy of Artem Basyrov, see Paper VII.

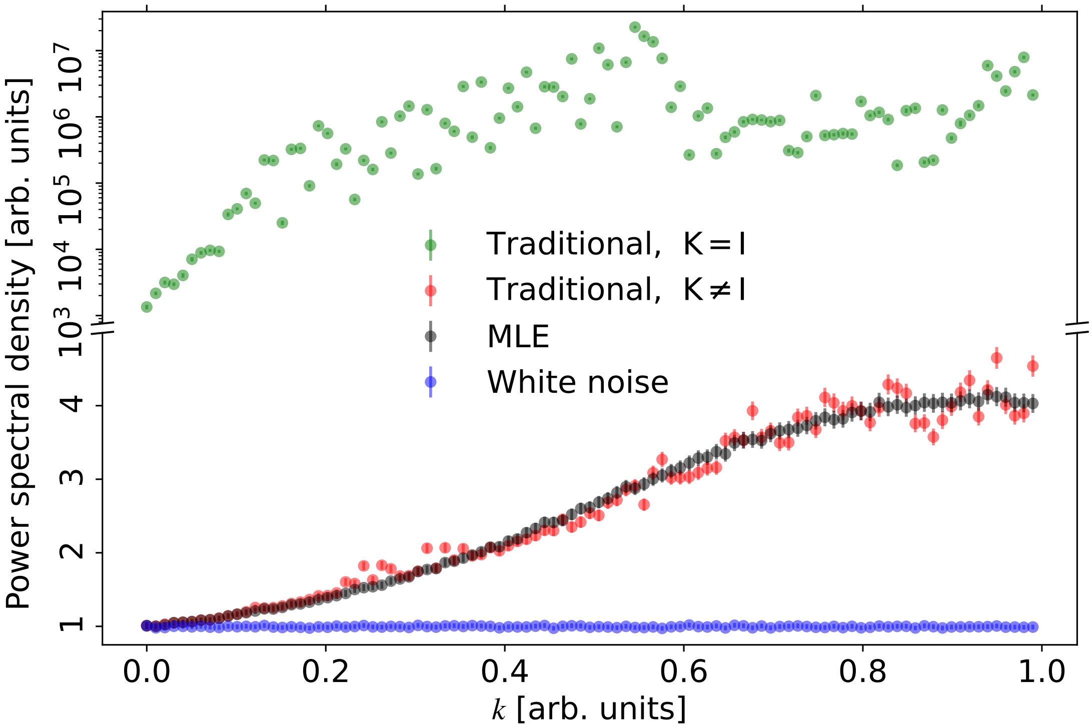


Figure 59: Mean 1D map power spectrum and corresponding 1σ uncertainties of 10,000 noise-only realizations for white noise (blue), the traditional Planck HFI deconvolution, with and without applying a low-pass filter K (respectively, red and green), as well as the MLE method of Paper VII. Note a broken linear and logarithmic y-axis. The figure is taken from Paper VII.

However, the resulting map $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ will be a biased signal estimator that will show significant signs of ringing along the survey scanning direction if the kernel $\mathsf{K} \neq \mathsf{I}$. Furthermore, the method does not use proper noise weights $\mathsf{K}\mathsf{T}^{-1}\mathsf{N}\mathsf{T}^{-1}\mathsf{K}$ in Eq. (4.18).

To solve this problem, we propose a maximum likelihood estimate (MLE) in Paper VII where the transfer function T is taken into account directly in the mapmaking stage. In doing so, we get the mapmaking equation

$$\mathsf{P}^T\mathsf{T}^T\mathsf{N}^{-1}\mathsf{T}\mathsf{P}\hat{\boldsymbol{m}}_{\mathrm{MLE}} = \mathsf{P}^T\mathsf{T}^T\mathsf{N}^{-1}\boldsymbol{d}, \tag{4.19}$$

where $\mathsf{T}^T = \mathsf{F}^{-1}\mathsf{T}^*(\omega)\mathsf{F}$. This equation must then be solved using conjugate gradient methods (see Shewchuk et al., 1994). As such, the method is significantly slower to apply than the traditional method employed by Planck HFI. The advantage is that the resulting MLE estimate of the map $\hat{\boldsymbol{m}}_{\mathrm{MLE}}$ will be an unbiased and properly weighted estimator of the signal. In the future, the MLE mapmaker will be integrated into the `Commander` Gibbs sampling framework that we explained earlier in Sec. 4.1.

To build some intuition on the proposed MLE mapmaker in Paper VII we apply it to a 1D toy model where the bolometer transfer function $T(\omega)$ and low-pass filter $K(\omega)$ used are identical to those of the Planck HFI 143-5 bolometer. Furthermore, we adopt a simple sinusoidal scanning strategy and apply the bolometer transfer

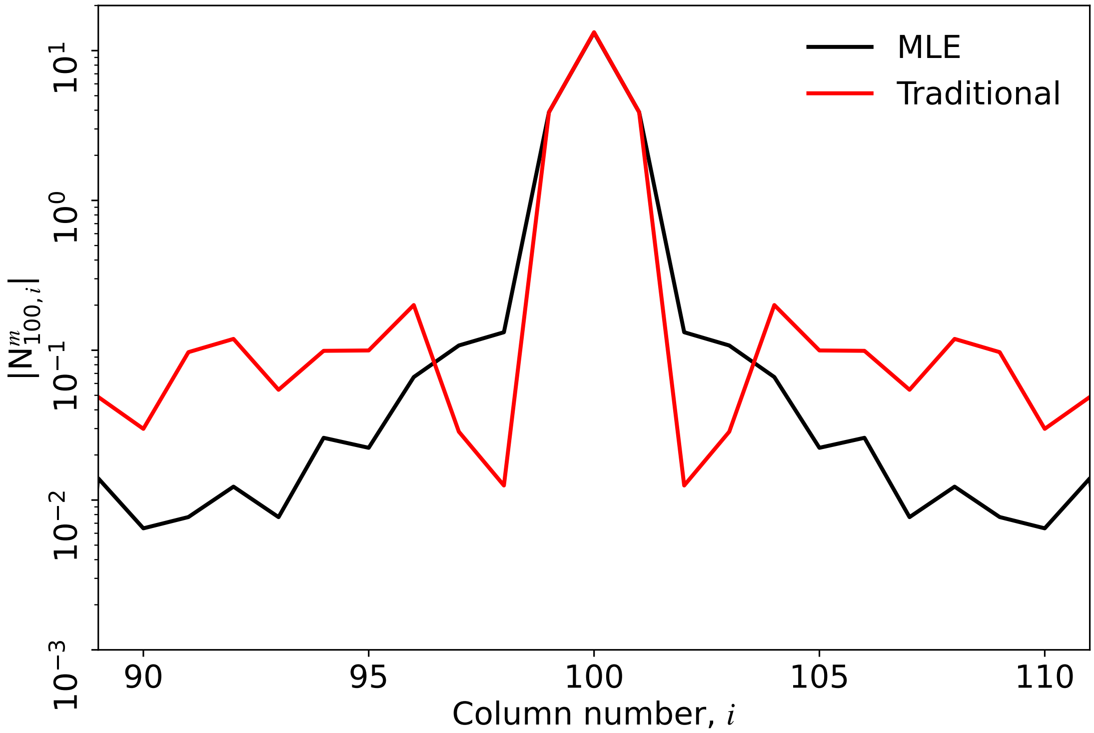


Figure 60: Absolute value of a slice through the covariance matrix, $|\mathsf{N}^m_{100,i}| = (\mathsf{P}^T N^{-1}\mathsf{P})^{-1}_{100,i}$, of the 1D MLE (black) and traditional (deconvolution and low-pass filtered) Planck HFI 1D toy map (red) seen in Fig. 59. The figure is taken from Paper VII.

function to the signal before adding noise. Subsequently, we generate 1D maps of 200 pixels using the traditional and proposed MLE methods described earlier. The result is seen in Fig. 59, which shows the average and standard deviation of the power spectrum of the toy model maps using the different algorithms. The traditional method dramatically increases noise without applying a low-pass filter $K(\omega)$. At the same time, it matches the MLE method noise level if low-pass filtering is used. Both the MLE and the traditional low-pass-filtered method provide power spectra above the white noise level of the data. However, we can also see that adjacent points in the $\mathsf{K} \neq \mathsf{I}$ case have a significantly longer correlation length than the MLE method. This is also reflected in Fig. 60 by the lower wings of a slice through the absolute ensemble covariance matrix of the MLE algorithm.

Furthermore, we see the effect of the HFI low-pass filter on a 1D signal-only run in Fig. 61, where we see that the input signal is reconstructed in an unbiased way by the MLE mapmaker. Meanwhile, the traditional method results in a bias that manifests as ringing in the fractional residual. This perfectly reflects what is seen in Fig. 62, where the traditional method shows asymmetric ringing in the residual along the scanning strategy. In contrast, the MLE residual appears as white noise.

In fact, as shown in Paper VII, employing the proposed MLE mapmaker, respectively, reduces the ellipticity and FWHM of the effective instrumental beam by 64 % and 2.3 %. For more details on the applications to the full Planck HFI

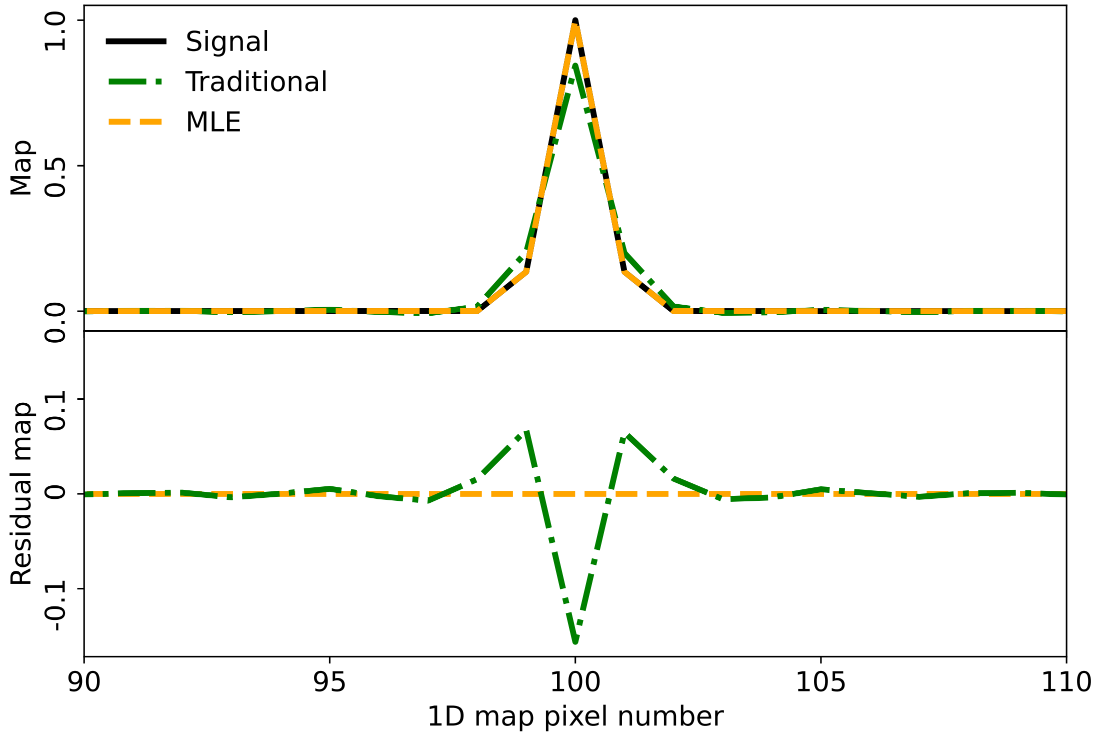


Figure 61: Simulated 1D point source (black), the reconstructions thereof, $\hat{\boldsymbol{m}}_{\mathrm{trad}}$, using the traditional Planck HFI deconvolution method (green). Meanwhile, the map $\hat{\boldsymbol{m}}_{\mathrm{MLE}}$ (orange) results from the proposed MLE algorithm of Paper VII. The lower panel shows the fractional residual between the input and output signal of the two methods. The figure is taken from Paper VII.

data, see Paper VII. Overall, we conclude that the proposed MLE mapmaker will be highly beneficial for current and future bolometric experiments, both in the CMB, LIM, or any other field.

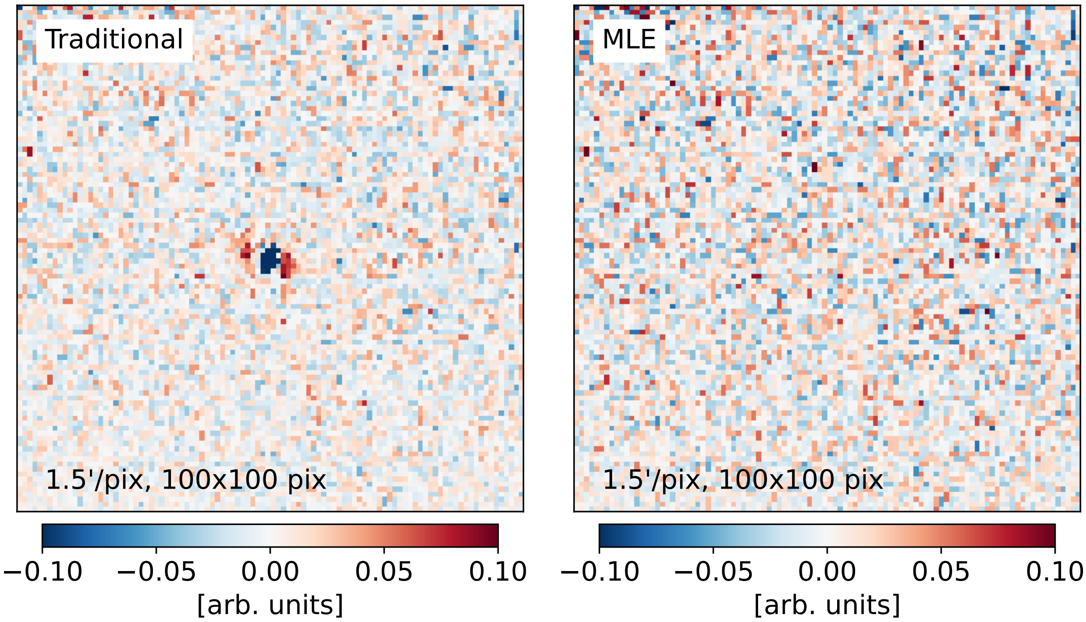


Figure 62: Residual of point source seen in Fig. 58 after estimating the signal using either (*left*) the traditional Planck HFI deconvolution method or (*right*) the proposed MLE method of Paper VII. Courtesy of Artem Basyrov, see Paper VII.

Chapter 5

Conclusion and future outlook

In this thesis, I developed and further improved a state-of-the-art data analysis pipeline for the COMAP-Pathfinder line intensity mapping experiment. COMAP aims to make the first large-scale maps of star-forming CO at the EoGA. Specifically, we have worked on two successive generations of the COMAP data analysis pipeline for the COMAP Early Science (Papers I and II and COMAP Season 2 releases (Papers III-V), respectively. Of these two, the most significant contribution of this work has been to the Season 2 analysis. Additionally, we have also worked on the `Commander3` global Bayesian Gibbs sampler as a future improvement for LIM experiments such as COMAP since it represents an algorithmically more well-motivated pipeline architecture compared to the current pipeline of COMAP (Papers VI and VII).

The Early Science and Season 2 pipelines have algorithmically similar architectures, with the latter version building and improving on the former. The raw data are captured with the COMAP-Pathfinder 26–34 GHz telescope at Owens Valley Radio Observatory that is sensitive to the CO(1–0) rotational transition of carbon monoxide emitted at redshifts $z \sim 2.4$–3.4 in the EoGA. Subsequently, as explained in Papers I and III, systematic contributions to the data from continuum foregrounds, $1/f$ gain fluctuations, ground pickup, standing waves in the telescope optics and electronics are removed from the data to isolate the extragalactic CO and white noise. The data are then calibrated and projected into sky maps. In the COMAP Season 2 release, Paper III, we discovered several new, highly pointing- and frequency-correlated systematic effects in the maps. These were coined as the start-of-scan and turn-around effects and are most likely sourced by standing waves in the telescope signal path due to mechanical vibrations as the instrument scans. We implemented a map-domain PCA filter to remove these systematic effects, which effectively removed their contribution to below the noise level by subtracting only 5 of 256 possible modes.

After filtering the maps with our map-PCA, we computed power spectra. Specifically, the COMAP power spectrum methodology described in Paper II, utilizes cross-power spectra between maps with data from different feeds and elevations. Each feed has independent noise and largely independent signal processing chains, and different elevations are expected to have various levels of ground pickup contamination. Thus, our feed-feed pseudo-cross-power spectra (FPXS) are both sensitive and robust against systematic effects.

In Paper IV, we improved our FPXS methodology by grouping detectors suspected to have similar levels of systematic effects and used these feed groups to compute cross-spectra (FGPXS). The resulting feed-group pseudo-cross-power spectra are somewhat more robust against feed-common systematic effects, as feeds with shared systematic effects are never cross-correlated. Furthermore, we improve the power spectrum uncertainty estimation using a data-driven approach that inherits all the correct properties of the real data. To ensure the data quality of our final power spectrum data products, we develop a large set of robust difference map null tests and find that they all pass according to noise expectations.

Using the COMAP ES pipeline of Paper I and II and about a year's worth of ES data, the resulting power spectrum represents the world's first direct 3D constraints on the CO(1–0) clustering power spectrum at the EoGA, ruling out several models from the literature. With additional data, the COMAP Season 2 raw data volume was about three times that of COMAP ES. Combined with an improved low-level analysis, the COMAP Season 2 analysis of Papers III and IV resulted in around an order of magnitude deeper upper limits on the CO(1–0) power spectrum than COMAP ES and COPSS (the only other CO(1–0) LIM experiment; Keating et al., 2016). As such, the COMAP Season 2 results present the currently tightest direct 3D CO(1–0) clustering constraints in the literature. The sensitivity is now large enough to further exclude the already excluded models of COMAP ES in *individual* k-bins and severely restricts the space of possible CO(1–0) models (Papers IV and V). As shown in Paper V, the COMAP Season 2 data favor the Li-Keating model. However, all non-excluded are consistent with the data to 2σ when assuming a simple two-parameter power spectrum model with a clustering and shot noise amplitude. Assuming a more empirically motivated five-parameter COMAP fiducial model, the COMAP Season 2 data shows the first hints of an increased faint end of the CO(1–0) luminosity function. As the COMAP data integrates down as noise, the analysis methods are still being refined, and observations continue; COMAP is expected to make a detection within the next couple of years. Our pipeline will be the first building block for future phases of COMAP.

In addition to working on the current COMAP pipeline, this thesis also considers some algorithmic improvements for a future iteration of a LIM analysis pipeline. Specifically, in Paper VI we showed the power of the `Commander3` Global Bayesian Gibbs sampler by mapping out the full posterior probability space of instrumental parameters in a set of simulated Planck LFI 30 GHz data. This served as a validation of key routines such as the mapmaking and gain and correlated noise estimation in the `Commander3` framework. The code worked as intended and provided a detailed overview of the posterior probability space with all corresponding non-trivial correlations and uncertainties of the system. Incorporating COMAP into `Commander3` will be a key future outlook beyond this thesis.

Lastly, in Paper VII we work on a maximum likelihood mapmaker to account for the Planck HFI bolometer transfer function directly in a conjugate gradient mapmaker, instead of relying on the deconvolution method of the original Planck HFI team. Specifically, we worked on validating the developed mapmaker using simplified 1D toy models of the Planck HFI 143 GHz data. The proposed maximum likelihood method resulted in a 64 % reduced ellipticity and a 2.3 % smaller full-width-at-half-maximum of the effective instrumental beam. Therefore, this method will be highly beneficial for any future CMB or LIM experiment that uses bolometric detectors.

Bibliography

Abazajian, K. N., et al. 2009, arXiv:0812.0649, ApJS, 182, 543, The Seventh Data Release of the Sloan Digital Sky Survey

Abdurashidova, Z., et al. 2022, arXiv:2108.02263, ApJ, 925, 221, First Results from HERA Phase I: Upper Limits on the Epoch of Reionization 21 cm Power Spectrum

Abdurro'uf, et al. 2022, arXiv:2112.02026, ApJS, 259, 35, The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data

Achúcarro, A., et al. 2022, arXiv:2203.08128, arXiv e-prints, arXiv:2203.08128, Inflation: Theory and Observations

Ade, P. A. R., et al. 2020, arXiv:1912.07118, Journal of Low Temperature Physics, 199, 1027, The Experiment for Cryogenic Large-Aperture Intensity Mapping (EXCLAIM)

Agrawal, S., Bracks, J., Fu, J., Janssen, R., Liu, L.-J., Lowe, I., Mayer, E., & Saeid, T. 2024, The Terahertz Intensity Mapper (TIM), LIM24 conference, Urbana-Champaign, Illinois, USA. https://mediaspace.illinois.edu/media/t/1_x5d61o2p

Alam, S., et al. 2021, arXiv:2007.08991, Phys. Rev. D, 103, 083533, Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory

Albrecht, A. & Steinhardt, P. J. 1982, Phys. Rev. Lett., 48, 1220, Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking

Amiri, M., et al. 2024, arXiv:2309.04404, ApJ, 963, 23, A Detection of Cosmological 21 cm Emission from CHIME in Cross-correlation with eBOSS Measurements of the Lyα Forest

Andersen, K. J., et al. 2023, arXiv:2201.08188, A&A, 675, A13, BEYONDPLANCK. XIII. Intensity foreground sampling, degeneracies, and priors

Anderson, C. J., et al. 2018, arXiv:1710.00424, MNRAS, 476, 3382, Low-amplitude clustering in low-redshift 21-cm intensity maps cross-correlated with 2dF galaxy densities

Annis, J., et al. 2014, arXiv:1111.6619, ApJ, 794, 120, The Sloan Digital Sky Survey Coadd: 275 deg^2 of Deep Sloan Digital Sky Survey Imaging on Stripe 82

Bardeen, J. M., Steinhardt, P. J., & Turner, M. S. 1983, Phys. Rev. D, 28, 679, Spontaneous creation of almost scale-free density perturbations in an inflationary universe

Barkana, R. & Loeb, A. 2001, astro-ph/0010468, Phys. Rep., 349, 125, In the beginning: the first sources of light and the reionization of the universe

Basyrov, A., Stutzer, N. O., Lunde, J. G. S., Eriksen, H. K., Gjerløw, E., Watts, D. J., & Wehus, I. K. 2024, A&A, 688, A166, Optimal bolometer transfer function deconvolution for CMB experiments through maximum likelihood mapmaking

Basyrov, A., et al. 2023, arXiv:2208.14293, A&A, 675, A10, BEYONDPLANCK. X. Planck Low Frequency Instrument frequency maps with sample-based error propagation

Battye, R. A., Davies, R. D., & Weller, J. 2004, astro-ph/0401340, MNRAS, 355, 1339, Neutral hydrogen surveys for high-redshift galaxy clusters and protoclusters

Becker, R. H., et al. 2001, astro-ph/0108097, AJ, 122, 2850, Evidence for Reionization at z~6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar

Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, arXiv:1806.07893, MNRAS, 488, 3143, UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10

Bennett, C. L., et al. 1994, astro-ph/9401012, ApJ, 436, 423, Cosmic Temperature Fluctuations from Two Years of COBE Differential Microwave Radiometers Observations

Bennett, C. L., et al. 1996, astro-ph/9601067, ApJ, 464, L1, Four-Year COBE DMR Cosmic Microwave Background Observations: Maps and Basic Results

Bennett, C. L., et al. 2003, astro-ph/0301158, ApJ, 583, 1, The Microwave Anisotropy Probe Mission

Bernal, J. L. & Kovetz, E. D. 2022, arXiv:2206.15377, A&A Rev., 30, 5, Line-intensity mapping: theory review with a focus on star-formation lines

Béthermin, M., et al. 2017, arXiv:1703.08795, A&A, 607, A89, The impact of clustering and angular resolution on far-infrared and millimeter continuum observations

Beutler, F., et al. 2011, arXiv:1106.3366, MNRAS, 416, 3017, The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant

BeyondPlanck Collaboration, et al. 2023, arXiv:2011.05609, A&A, 675, A1, BEYONDPLANCK. I. Global Bayesian analysis of the Planck Low Frequency Instrument data

Blake, C., et al. 2012, arXiv:1204.3674, MNRAS, 425, 405, The WiggleZ Dark Energy Survey: joint measurements of the expansion and growth history at $z < 1$

Bodenheimer, P. H. 2011, Principles of Star Formation, 1st edn. (Springer Berlin)

Bond, J. R. & Myers, S. T. 1996, ApJS, 103, 1, The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Caruana, J., Holwerda, B., Smit, R., & Wilkins, S. 2015, arXiv:1503.08228, ApJ, 811, 140, Reionization After Planck: The Derived Growth of the Cosmic Ionizing Emissivity Now Matches the Growth of the Galaxy UV Luminosity Density

Breysse, P. C., Anderson, C. J., & Berger, P. 2019, arXiv:1907.04369, Phys. Rev. Lett., 123, 231105, Canceling Out Intensity Mapping Foregrounds

Breysse, P. C., Chung, D. T., & Ihle, H. T. 2023, arXiv:2210.14902, MNRAS, 525, 1824, Characteristic functions for cosmological cross-correlations

Breysse, P. C., Kovetz, E. D., Behroozi, P. S., Dai, L., & Kamionkowski, M. 2017, arXiv:1609.01728, MNRAS, 467, 2996, Insights from probability distribution functions of intensity maps

Breysse, P. C., Kovetz, E. D., & Kamionkowski, M. 2015, arXiv:1503.05202, MNRAS, 452, 3408, Masking line foregrounds in intensity-mapping surveys

Breysse, P. C., et al. 2022, arXiv:2111.05933, ApJ, 933, 188, COMAP Early Science. VII. Prospects for CO Intensity Mapping at Reionization

Brilenkov, M., et al. 2023, arXiv:2209.04437, A&A, 675, A4, BEYONDPLANCK. IV. Simulations and validation

Carroll, S. M. 2019, Spacetime and Geometry: An Introduction to General Relativity (Cambridge University Press)

CCAT-Prime Collaboration, et al. 2023, arXiv:2107.10364, ApJS, 264, 7, CCAT-prime Collaboration: Science Goals and Forecasts with Prime-Cam on the Fred Young Submillimeter Telescope

Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Nature, 466, 463, An intensity map of hydrogen 21-cm emission at redshift z$\sim$0.8

Cheng, Y.-T., Chang, T.-C., Bock, J., Bradford, C. M., & Cooray, A. 2016, arXiv:1604.07833, ApJ, 832, 165, Spectral Line De-confusion in an Intensity Mapping Survey

CHIME Collaboration, et al. 2022, arXiv:2201.07869, ApJS, 261, 29, An Overview of CHIME, the Canadian Hydrogen Intensity Mapping Experiment

Chung, D. T., et al. 2019, arXiv:1809.04550, ApJ, 872, 186, Cross-correlating Carbon Monoxide Line-intensity Maps with Spectroscopic and Photometric Galaxy Surveys

Chung, D. T., et al. 2021, arXiv:2104.11171, ApJ, 923, 188, A Model of Spectral Line Broadening in Signal Forecasts for Line-intensity Mapping Experiments

Chung, D. T., et al. 2022a, arXiv:2111.05931, ApJ, 933, 186, COMAP Early Science. V. Constraints and Forecasts at z 3

Chung, D. T., et al. 2023, arXiv:2210.14890, MNRAS, 520, 5305, The deconvolved distribution estimator: enhancing reionization-era CO line-intensity mapping analyses with a cross-correlation analogue for one-point statistics

Chung, D. T., et al. 2024, arXiv:2406.07512, A&A, 691, A337, COMAP Pathfinder – Season 2 results: III. Implications for cosmic molecular gas content at z∼3

Chung, T. D. 2022b, PhD thesis, Stanford University, lINE-INTENSITY MAPPING WITH THE CO MAPPING ARRAY PATHFINDER AND BEYOND, https://purl.stanford.edu/yv096yc6652

Cleary, K. A., et al. 2022, arXiv:2111.05927, ApJ, 933, 182, COMAP Early Science. I. Overview

Cochrane, R. K., et al. 2023, arXiv:2305.15510, MNRAS, 523, 6082, The LOFAR Two-metre Sky Survey: the radio view of the cosmic star formation history

Cole, S., et al. 2005, astro-ph/0501174, MNRAS, 362, 505, The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications

Colless, M., et al. 2001, astro-ph/0106498, MNRAS, 328, 1039, The 2dF Galaxy Redshift Survey: spectra and redshifts

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., & Broderick, J. J. 2002, VizieR Online Data Catalog: 1.4GHz NRAO VLA Sky Survey (NVSS) (Condon+ 1998), VizieR On-line Data Catalog: VIII/65. Originally published in: 1998AJ....115.1693C

Crichton, D., et al. 2022, arXiv:2109.13755, Journal of Astronomical Telescopes, Instruments, and Systems, 8, 011019, Hydrogen Intensity and Real-Time Analysis Experiment: 256-element array status and overview

Crites, A. & Lau, K. 2024, Probing EoR with TIME: overview, update and commissioning observations, LIM24 conference, Urbana-Champaign, Illinois, USA. https://mediaspace.illinois.edu/media/t/1_bjj3siqz

Cunnington, S., Irfan, M. O., Carucci, I. P., Pourtsidou, A., & Bobin, J. 2021, arXiv:2010.02907, MNRAS, 504, 208, 21-cm foregrounds and polarization leakage: cleaning and mitigation strategies

Cunnington, S., et al. 2023, arXiv:2206.01579, MNRAS, 518, 6262, H I intensity mapping with MeerKAT: power spectrum detection in cross-correlation with WiggleZ galaxies

Demtroder, W. 2010, Atoms, Molecules and Photons: An Introduction to Atomic-, Molecular- and Quantum Physics, 2nd edn., Graduate Texts in Physics (Berlin, Heidelberg: Springer Nature)

Dodelson, S. 2003, Modern Cosmology

Doré, O., et al. 2018, arXiv:1805.05489, arXiv e-prints, arXiv:1805.05489, Science Impacts of the SPHEREx All-Sky Optical to Near-Infrared Spectral Survey II: Report of a Community Workshop on the Scientific Synergies Between the SPHEREx Survey and Other Astronomy Observatories

Drinkwater, M. J., et al. 2010, arXiv:0911.4246, MNRAS, 401, 1429, The WiggleZ Dark Energy Survey: survey design and first data release

Dunne, D. A., et al. 2024, arXiv:2304.09832, ApJ, 965, 7, COMAP Early Science. VIII. A Joint Stacking Analysis with eBOSS Quasars

Dunne, D. A., et al. 2025, arXiv:2503.21743, Submitted to A&A, Three-Dimensional Stacking as a Line Intensity Mapping Statistic

Einstein, A. 1915, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 844, Die Feldgleichungen der Gravitation

Eisenstein, D. J., et al. 2005, astro-ph/0501171, ApJ, 633, 560, Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies

Eriksen, H. K., Jewell, J. B., Dickinson, C., Banday, A. J., Górski, K. M., & Lawrence, C. R. 2008, arXiv:0709.1058, ApJ, 676, 10, Joint Bayesian Component Separation and CMB Power Spectrum Estimation

Eriksen, H. K., et al. 2004, astro-ph/0407028, ApJS, 155, 227, Power Spectrum Estimation from High-Resolution Maps by Gibbs Sampling

Eskilt, J. R., et al. 2023, arXiv:2305.02268, A&A, 679, A144, COSMOGLOBE DR1 results. II. Constraints on isotropic cosmic birefringence from reprocessed WMAP and Planck LFI data

Eskilt, J. R., et al. 2023a, arXiv:2306.15511, A&A, 678, A169, COSMOGLOBE: Towards end-to-end CMB cosmological parameter estimation without likelihood approximations

Eskilt, J. R., et al. 2023b, arXiv:2305.02268, A&A, 679, A144, COSMOGLOBE DR1 results. II. Constraints on isotropic cosmic birefringence from reprocessed WMAP and Planck LFI data

Fasano, A., et al. 2024, in European Physical Journal Web of Conferences, Vol. 293, mm Universe 2023 - Observing the Universe at mm Wavelengths, 00018

Fixsen, D. J., Cheng, E. S., Gales, J. M., Mather, J. C., Shafer, R. A., & Wright, E. L. 1996, astro-ph/9605054, ApJ, 473, 576, The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set

Foss, M. K., et al. 2022, arXiv:2111.05929, ApJ, 933, 184, COMAP Early Science. III. CO Data Processing

Fronenberg, H. & Liu, A. 2024, arXiv:2407.14588, ApJ, 975, 222, Forecasts and Statistical Insights for Line Intensity Mapping Cross-correlations: A Case Study with 21 cm × [C II]

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, astro-ph/0608032, Phys. Rep., 433, 181, Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe

Fuskeland, U., et al. 2023, arXiv:2302.05228, A&A, 676, A42, Tensor-to-scalar ratio forecasts for extended LiteBIRD frequency configurations

Galloway, M., et al. 2023, arXiv:2201.03478, A&A, 675, A8, BEYONDPLANCK. VIII. Efficient sidelobe convolution and corrections through spin harmonics

Gaunt, R. E. 2019, https://onlinelibrary.wiley.com/doi/pdf/10.1111/stan.12152, Statistica Neerlandica, 73, 176, A note on the distribution of the product of zero-mean correlated normal random variables, https://onlinelibrary.wiley.com/doi/abs/10.1111/stan.12152

Gebhardt, K., et al. 2021, arXiv:2110.04298, ApJ, 923, 217, The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) Survey Design, Reductions, and Detections

Geman, S. & Geman, D. 1984, IEEE Transactions on pattern analysis and machine intelligence, 721, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images

Gerakakis, S., et al. 2023, arXiv:2205.11262, The Open Journal of Astrophysics, 6, 10, From BeyondPlanck to Cosmoglobe: Open Science, Reproducibility, and Data Longevity

Gjerløw, E., et al. 2023, arXiv:2011.08082, A&A, 675, A7, BEYONDPLANCK. VII. Bayesian estimation of gain and absolute calibration for cosmic microwave background experiments

Gong, Y., Cooray, A., Silva, M., Santos, M. G., Bock, J., Bradford, C. M., & Zemcov, M. 2012, arXiv:1107.3553, ApJ, 745, 49, Intensity Mapping of the [C II] Fine Structure Line during the Epoch of Reionization

Gong, Y., Silva, M., Cooray, A., & Santos, M. G. 2014, arXiv:1312.2035, ApJ, 785, 72, Foreground Contamination in Lyα Intensity Mapping during the Epoch of Reionization

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, astro-ph/0409513, ApJ, 622, 759, HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere

Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. 2022, Nature Reviews Methods Primers, 2, 100, Principal component analysis

Gruppioni, C., et al. 2013, arXiv:1302.5209, MNRAS, 432, 23, The Herschel PEP/HerMES luminosity function - I. Probing the evolution of PACS selected Galaxies to z = 4

Gunn, J. E. & Peterson, B. A. 1965, ApJ, 142, 1633, On the Density of Neutral Hydrogen in Intergalactic Space.

Guth, A. H. 1981, Phys. Rev. D, 23, 347, Inflationary universe: A possible solution to the horizon and flatness problems

Guth, A. H. & Pi, S. Y. 1982, Phys. Rev. Lett., 49, 1110, Fluctuations in the New Inflationary Universe

Hamilton, A. J. S. 1998, astro-ph/9708102, in Astrophysics and Space Science Library, Vol. 231, The Evolving Universe, ed. D. Hamilton, 185

Harper, S. E., et al. 2025, arXiv:2405.04383, MNRAS, 536, 2914, COMAP Galactic science I: observations of spinning dust emission at 30 GHz in dark clouds surrounding the λ-Orionis H II region

Hashimoto, T., et al. 2018, arXiv:1805.05966, Nature, 557, 392, The onset of star formation 250 million years after the Big Bang

Hawking, S. W. 1982, Physics Letters B, 115, 295, The development of irregularities in a single bubble inflationary universe

HERA Collaboration, et al. 2023, arXiv:2210.04912, ApJ, 945, 124, Improved Constraints on the 21 cm EoR Power Spectrum and the X-Ray Heating of the IGM with HERA Phase I Observations

Hill, R., Masui, K. W., & Scott, D. 2018, arXiv:1802.03694, Applied Spectroscopy, 72, 663, The Spectrum of the Universe

Hivon, E., Górski, K. M., Netterfield, C. B., Crill, B. P., Prunet, S., & Hansen, F. 2002, astro-ph/0105302, ApJ, 567, 2, MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets

Hollenbach, D. J. & Tielens, A. G. G. M. 1999, Reviews of Modern Physics, 71, 173, Photodissociation regions in the interstellar medium of galaxies

Hu, W. & White, M. 1997, astro-ph/9706147, New A, 2, 323, A CMB polarization primer

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168, A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae

Huynh, M. T., et al. 2017, arXiv:1701.05698, MNRAS, 467, 1222, The AT-LESS CO(1-0) survey of submillimetre galaxies in the Extended Chandra Deep Field South: First results on cold molecular gas in galaxies at z $\sim$ 2

Ihle, H. T., et al. 2019, arXiv:1808.07487, ApJ, 871, 75, Joint Power Spectrum and Voxel Intensity Distribution Forecast on the CO Luminosity Function with COMAP

Ihle, H. T., et al. 2022, arXiv:2111.05930, ApJ, 933, 185, COMAP Early Science. IV. Power Spectrum Methodology and Results

Ihle, H. T., et al. 2023, arXiv:2011.06650, A&A, 675, A6, BEYONDPLANCK. VI. Noise characterization and modeling

Jia, W., Sun, M., Lian, J., & Hou, S. 2022, Complex & Intelligent Systems, 8, 2663, Feature dimensionality reduction: a review

Jones, D. H., et al. 2009, arXiv:0903.5451, MNRAS, 399, 683, The 6dF Galaxy Survey: final redshift release (DR3) and southern large-scale structures

Kaiser, N. 1987, MNRAS, 227, 1, Clustering in real space and in redshift space

Kamionkowski, M., Kosowsky, A., & Stebbins, A. 1997, astro-ph/9609132, Phys. Rev. Lett., 78, 2058, A Probe of Primordial Gravity Waves and Vorticity

Kamionkowski, M. & Kovetz, E. D. 2016, arXiv:1510.06042, ARA&A, 54, 227, The Quest for B Modes from Inflationary Gravitational Waves

Kamionkowski, M. & Riess, A. G. 2023, arXiv:2211.04492, Annual Review of Nuclear and Particle Science, 73, 153, The Hubble Tension and Early Dark Energy

Karkare, K., et al. 2022, Journal of Low Temperature Physics, 209, 758, SPT-SLIM: a line intensity mapping pathfinder for the south pole telescope

Kazanas, D. 1980, ApJ, 241, L59, Dynamics of the universe and spontaneous symmetry breaking

Keating, G. K., Marrone, D. P., Bower, G. C., & Keenan, R. P. 2020, arXiv:2008.08087, ApJ, 901, 141, An Intensity Mapping Detection of Aggregate CO Line Emission at 3 mm

Keating, G. K., Marrone, D. P., Bower, G. C., Leitch, E., Carlstrom, J. E., & DeBoer, D. R. 2016, arXiv:1605.03971, ApJ, 830, 34, COPSS II: The Molecular Gas Content of Ten Million Cubic Megaparsecs at Redshift z $\sim$ 3

Keating, G. K., et al. 2015, arXiv:1510.06744, ApJ, 814, 140, First Results from COPSS: The CO Power Spectrum Survey

Keihänen, E., et al. 2023, arXiv:2011.06024, A&A, 675, A2, BEYONDPLANCK. II. CMB mapmaking through Gibbs sampling

Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, arXiv:1910.09730, ARA&A, 57, 511, Understanding Galaxy Evolution Through Emission Lines

Knox, L. 1995, astro-ph/9504054, Phys. Rev. D, 52, 4307, Determination of inflationary observables by cosmic microwave background anisotropy experiments

Kofman, L., Linde, A., & Starobinsky, A. A. 1994, hep-th/9405187, Phys. Rev. Lett., 73, 3195, Reheating after inflation

Komatsu, E., et al. 2009, arXiv:0803.0547, ApJS, 180, 330, Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation

Kovetz, E. D., et al. 2017, arXiv:1709.09066, arXiv e-prints, arXiv:1709.09066, Line-Intensity Mapping: 2017 Status Report

Lagache, G., Cousin, M., & Chatzikos, M. 2018, arXiv:1711.00798, A&A, 609, A130, The [CII] 158 μm line emission in high-redshift galaxies

Lagache, G., Puget, J.-L., & Dole, H. 2005, astro-ph/0507298, ARA&A, 43, 727, Dusty Infrared Galaxies: Sources of the Cosmic Infrared Background

Lamb, J. W., et al. 2022, arXiv:2111.05928, ApJ, 933, 183, COMAP Early Science. II. Pathfinder Instrument

Leighton, R. B. 1977, A 10-Meter Telescope for Millimeter and Sub-Millimeter Astronomy, Final Technical Report for National Science Foundation Research Grant AST 73-04908, January 1974 - August 1977, 87 pages

Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49, Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques

Leung, J. S. Y., et al. 2022, arXiv:2111.01113, ApJ, 928, 109, A Simulation-based Method for Correcting Mode Coupling in CMB Angular Power Spectra

Li, T. Y., Wechsler, R. H., Devaraj, K., & Church, S. E. 2016, arXiv:1503.08833, ApJ, 817, 169, Connecting CO Intensity Mapping to Molecular Gas and Star Formation in the Epoch of Galaxy Assembly

Lidz, A. & Taylor, J. 2016, arXiv:1604.05737, ApJ, 825, 143, On Removing Interloper Contamination from Intensity Mapping Power Spectrum Measurements

Linde, A. D. 1982, Physics Letters B, 108, 389, A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems

Loeb, A. & Wyithe, J. S. B. 2008, arXiv:0801.1677, Phys. Rev. Lett., 100, 161301, Possibility of Precise Measurement of the Cosmological Power Spectrum with a Dedicated Survey of 21cm Emission after Reionization

Lunde, J. G. S., et al. 2024, arXiv:2406.07510, A&A, 691, A335, COMAP Pathfinder – Season 2 results: I. Improved data selection and processing

Madau, P. & Dickinson, M. 2014, arXiv:1403.0007, ARA&A, 52, 415, Cosmic Star-Formation History

Madau, P., Meiksin, A., & Rees, M. J. 1997, astro-ph/9608010, ApJ, 475, 429, 21 Centimeter Tomography of the Intergalactic Medium at High Redshift

Mashian, N., Loeb, A., & Sternberg, A. 2016, arXiv:1601.02618, MNRAS, 458, L99, Spectral distortion of the CMB by the cumulative CO emission from galaxies throughout cosmic history

Masui, K. W., et al. 2013, arXiv:1208.0331, ApJ, 763, L20, Measurement of 21 cm Brightness Fluctuations at z $\sim$0.8 in Cross-correlation

Mather, J. C., et al. 1990, ApJ, 354, L37, A Preliminary Measurement of the Cosmic Microwave Background Spectrum by the Cosmic Background Explorer (COBE) Satellite

Mather, J. C., et al. 1994, ApJ, 420, 439, Measurement of the Cosmic Microwave Background Spectrum by the COBE FIRAS Instrument

Mauch, T., et al. 2020, arXiv:1912.06212, ApJ, 888, 61, The 1.28 GHz MeerKAT DEEP2 Image

Morales, M. F. & Wyithe, J. S. B. 2010, arXiv:0910.3010, ARA&A, 48, 127, Reionization and Cosmology with 21-cm Fluctuations

Mukhanov, V. F. 1985, Soviet Journal of Experimental and Theoretical Physics Letters, 41, 493, Gravitational instability of the universe filled with a scalar field

Mukhanov, V. F. & Chibisov, G. V. 1981, ZhETF Pisma Redaktsiiu, 33, 549, Quantum fluctuations and a nonsingular universe

Mukhanov, V. F. & Chibisov, G. V. 1982, Soviet Journal of Experimental and Theoretical Physics, 56, 258, Vacuum energy and large-scale structure of the Universe

Nadarajah, S. & Pogány, T. K. 2016, Comptes Rendus. Mathématique, 354, 201, On the distribution of the product of correlated normal random variables

O'Dwyer, I. J., et al. 2004, astro-ph/0407027, ApJ, 617, L99, Bayesian Power Spectrum Analysis of the First-Year Wilkinson Microwave Anisotropy Probe Data

Oxholm, T. M. & Switzer, E. R. 2021, arXiv:2107.02111, Phys. Rev. D, 104, 083501, Intensity mapping without cosmic variance

Padmanabhan, H. 2018, arXiv:1706.01471, MNRAS, 475, 1477, Constraining the CO intensity mapping power spectrum at intermediate redshifts

Papovich, C., et al. 2016, arXiv:1603.05660, ApJS, 224, 28, The Spitzer-HETDEX Exploratory Large-area Survey

Particle Data Group Collaboration. 2022, PTEP, 2022, 083C01, Review of Particle Physics

Paul, S., Santos, M. G., Chen, Z., & Wolz, L. 2023, arXiv:2301.11943, arXiv e-prints, arXiv:2301.11943, A first detection of neutral hydrogen intensity mapping on Mpc scales at $z \approx 0.32$ and $z \approx 0.44$

Pavesi, R., et al. 2018, arXiv:1808.04372, ApJ, 864, 49, The CO Luminosity Density at High-z (COLDz) Survey: A Sensitive, Large-area Blind Search for Low-J CO Emission from Cold Gas in the Early Universe with the Karl G. Jansky Very Large Array

Pearson, K. 1901, The London, Edinburgh, and Dublin philosophical magazine and journal of science, 2, 559, LIII. On lines and planes of closest fit to systems of points in space

Penzias, A. A. & Burrus, C. A. 1973, ARA&A, 11, 51, Millimeter-Wavelength Radio-Astronomy Techniques

Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419, A Measurement of Excess Antenna Temperature at 4080 Mc/s.

Perlmutter, S., et al. 1999, astro-ph/9812133, ApJ, 517, 565, Measurements of Ω and Λ from 42 High-Redshift Supernovae

Peterson, J. B., Bandura, K., & Pen, U. L. 2006, astro-ph/0606104, arXiv e-prints, astro–ph/0606104, The Hubble Sphere Hydrogen Survey

Pillepich, A., et al. 2018, arXiv:1703.02970, MNRAS, 473, 4077, Simulating galaxy formation with the IllustrisTNG model

Planck Collaboration I. 2011, arXiv:1101.2022, A&A, 536, A1, Planck early results. I. The Planck mission

Planck Collaboration I. 2014, arXiv:1303.5062, A&A, 571, A1, Planck 2013 results. I. Overview of products and scientific results

Planck Collaboration I. 2020, arXiv:1807.06205, A&A, 641, A1, Planck 2018 results. I. Overview and the cosmological legacy of Planck

Planck Collaboration II. 2020, arXiv:1807.06206, A&A, 641, A2, Planck 2018 results. II. Low Frequency Instrument data processing

Planck Collaboration IV. 2020, arXiv:1807.06208, A&A, 641, A4, Planck 2018 results. IV. Diffuse component separation

Planck Collaboration IX. 2020, arXiv:1905.05697, A&A, 641, A9, Planck 2018 results. IX. Constraints on primordial non-Gaussianity

Planck Collaboration V. 2020, arXiv:1907.12875, A&A, 641, A5, Planck 2018 results. V. CMB power spectra and likelihoods

Planck Collaboration VI. 2020, arXiv:1807.06209, A&A, 641, A6, Planck 2018 results. VI. Cosmological parameters

Planck Collaboration VII. 2014, arXiv:1303.5068, A&A, 571, A7, *Planck* 2013 results. VII. HFI time response and beams

Planck Collaboration VII. 2016, arXiv:1502.01586, A&A, 594, A7, *Planck* 2015 results. VII. High Frequency Instrument data processing: Time-ordered information and beam processing

Planck Collaboration X. 2016, arXiv:1502.01588, A&A, 594, A10, Planck 2015 results. X. Diffuse component separation: Foreground maps

Planck Collaboration X. 2020, arXiv:1807.06211, A&A, 641, A10, Planck 2018 results. X. Constraints on inflation

Planck Collaboration XXX. 2014, arXiv:1309.0382, A&A, 571, A30, Planck 2013 results. XXX. Cosmic infrared background measurements and implications for star formation

Pullen, A. R., Chang, T.-C., Doré, O., & Lidz, A. 2013, ApJ, 768, 15, Cross-correlations as a cosmological carbon monoxide detector

Rahimi, M. & Reichardt, C. L. 2024, arXiv:2412.04099, arXiv e-prints, arXiv:2412.04099, Polarization of the Cosmic Microwave Background

Rennie, T. J., et al. 2022, arXiv:2111.05932, ApJ, 933, 187, COMAP Early Science. VI. A First Look at the COMAP Galactic Plane Survey

Richards, P. L. 1994, Journal of Applied Physics, 76, 1, Bolometers for infrared and millimeter waves

Riechers, D. A., et al. 2019, arXiv:1808.04371, ApJ, 872, 7, COLDz: Shape of the CO Luminosity Function at High Redshift and the Cold Gas History of the Universe

Riess, A. G., et al. 1998, astro-ph/9805201, AJ, 116, 1009, Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant

Riess, A. G., et al. 2022a, arXiv:2112.04510, ApJ, 934, L7, A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km s^{-1} Mpc^{-1} Uncertainty from the Hubble Space Telescope and the SH0ES Team

Riess, A. G., et al. 2022b, arXiv:2208.01045, ApJ, 938, 36, Cluster Cepheids with High Precision Gaia Parallaxes, Low Zero-point Uncertainties, and Hubble Space Telescope Photometry

Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, arXiv:1502.02024, ApJ, 802, L19, Cosmic Reionization and Early Star-forming Galaxies: A Joint Analysis of New Constraints from Planck and the Hubble Space Telescope

Robertson, B. E., et al. 2013, arXiv:1301.1228, ApJ, 768, 71, New Constraints on Cosmic Reionization from the 2012 Hubble Ultra Deep Field Campaign

Rowan-Robinson, M., et al. 2016, arXiv:1605.03937, MNRAS, 461, 1100, The star formation rate density from z = 1 to 6

Roy, A., Battaglia, N., & Pullen, A. R. 2024, arXiv:2406.07861, arXiv e-prints, arXiv:2406.07861, A Measurement of CO(3-2) Line Emission from eBOSS Galaxies at $z \sim 0.5$ using Planck Data

San, M., et al. 2024, arXiv:2408.11004, arXiv e-prints, arXiv:2408.11004, Cosmoglobe DR2. III. Improved modelling of zodiacal light with COBE-DIRBE through global Bayesian analysis

Santos, M. G., et al. 2015, arXiv preprint arXiv:1501.03989, Cosmology with a SKA HI intensity mapping survey

Sato, K. 1981, MNRAS, 195, 467, First-order phase transition of a vacuum and the expansion of the Universe

Schneider, P. 2015, Extragalactic Astronomy and Cosmology: An Introduction, 2nd edn. (Springer Berlin)

Schulz, N. S. 2012, The Formation and Early Evolution of Stars, 2nd edn. (Springer Berlin)

Seljak, U. & Zaldarriaga, M. 1997, astro-ph/9609169, Phys. Rev. Lett., 78, 2054, Signature of Gravity Waves in the Polarization of the Microwave Background

Seljebotn, D. S., Bærland, T., Eriksen, H. K., Mardal, K. A., & Wehus, I. K. 2019, A&A, 627, A98, Multi-resolution Bayesian CMB component separation through Wiener filtering with a pseudo-inverse preconditioner

Shewchuk, J. R. et al. 1994, An introduction to the conjugate gradient method without the agonizing pain

Shimabukuro, H., Yoshiura, S., Takahashi, K., Yokoyama, S., & Ichiki, K. 2016, arXiv:1507.01335, MNRAS, 458, 3003, 21 cm line bispectrum as a method to probe cosmic dawn and epoch of reionization

Silva, M., Santos, M. G., Cooray, A., & Gong, Y. 2015, arXiv:1410.4808, ApJ, 806, 209, Prospects for Detecting C II Emission during the Epoch of Reionization

Silva, M. B., Santos, M. G., Gong, Y., Cooray, A., & Bock, J. 2013, arXiv:1205.1493, ApJ, 763, 132, Intensity Mapping of Lyα Emission during the Epoch of Reionization

Silva, M. B., et al. 2021, arXiv:2111.05354, arXiv e-prints, arXiv:2111.05354, Synergies between the COMAP CO Line Intensity Mapping mission and a Lyα galaxy survey: How to probe the early universe with voxel based analysis of observational data

Smoot, G. F., et al. 1992, ApJ, 396, L1, Structure in the COBE Differential Microwave Radiometer First-Year Maps

Staniszewski, Z., et al. 2014, Journal of low temperature physics, 176, 767, The Tomographic Ionized-Carbon Mapping Experiment (TIME) CII Imaging Spectrometer

Starobinsky, A. A. 1982, Physics Letters B, 117, 175, Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations

Stein, G., Alvarez, M. A., & Bond, J. R. 2019, arXiv:1810.07727, MNRAS, 483, 2236, The mass-Peak Patch algorithm for fast generation of deep all-sky dark matter halo catalogues and its N-body validation

Stein, G., Alvarez, M. A., Bond, J. R., van Engelen, A., & Battaglia, N. 2020, arXiv:2001.08787, J. Cosmology Astropart. Phys., 2020, 012, The Websky extragalactic CMB simulations

Stutzer, N. O., et al. 2024, arXiv:2406.07511, A&A, 691, A336, COMAP Pathfinder – Season 2 results: II. Updated constraints on the CO(1–0) power spectrum

Sun, G., et al. 2018, arXiv:1610.10095, ApJ, 856, 107, A Foreground Masking Strategy for [C II] Intensity Mapping Experiments Using Galaxies Selected by Stellar Mass and Redshift

Svalheim, T. L., et al. 2023, arXiv:2011.08503, A&A, 675, A14, BEYOND-PLANCK. XIV. Polarized foreground emission between 30 and 70 GHz

Switzer, E. R., et al. 2021, Journal of Astronomical Telescopes, Instruments, and Systems, 7, 044004, Experiment for cryogenic large-aperture intensity mapping: instrument design

Tegmark, M. 1997, astro-ph/9611130, ApJ, 480, L87, How to Make Maps from Cosmic Microwave Background Data without Losing Information

Van Cuyck, M., et al. 2023, arXiv:2306.01568, A&A, 676, A62, CONCERTO: Extracting the power spectrum of the [C_{II}] emission line

van Haarlem, M. P., et al. 2013, Astronomy & astrophysics, 556, A2, LOFAR: The low-frequency array

Vazquez, J. A., Padilla, L. E., & Matos, T. 2018, arXiv:1810.09934, arXiv e-prints, arXiv:1810.09934, Inflationary Cosmology: From Theory to Observations

Vieira, J., et al. 2020, arXiv:2009.14340, arXiv e-prints, arXiv:2009.14340, The Terahertz Intensity Mapper (TIM): a Next-Generation Experiment for Galaxy Evolution Studies

Visbal, E. & Loeb, A. 2010, arXiv:1008.3178, J. Cosmology Astropart. Phys., 2010, 016, Measuring the 3D clustering of undetected galaxies through cross correlation of their cumulative flux fluctuations from multiple spectral lines

Visbal, E., Trac, H., & Loeb, A. 2011, arXiv:1104.4809, J. Cosmology Astropart. Phys., 2011, 010, Demonstrating the feasibility of line intensity mapping using mock data of galaxy clustering from simulations

Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, astro-ph/0310080, Phys. Rev. D, 70, 083511, Global, exact cosmic microwave background data analysis using Gibbs sampling

Watts, D. J., Addison, G. E., Bennett, C. L., & Weiland, J. L. 2020, arXiv:1910.00590, ApJ, 889, 130, Beyond Optical Depth: Future Determination of Ionization History from the Cosmic Microwave Background

Watts, D. J., et al. 2023a, arXiv:2202.11979, A&A, 675, A16, From BEYOND-PLANCK to COSMOGLOBE: Preliminary WMAP Q-band analysis

Watts, D. J., et al. 2023b, arXiv:2303.08095, A&A, 679, A143, COSMOGLOBE DR1 results. I. Improved Wilkinson Microwave Anisotropy Probe maps through Bayesian end-to-end analysis

Watts, D. J., et al. 2024a, arXiv:2406.01491, arXiv e-prints, arXiv:2406.01491, Cosmoglobe DR2. II. CIB monopole measurements from COBE-DIRBE through global Bayesian analysis

Watts, D. J., et al. 2024b, arXiv:2310.13740, A&A, 686, A297, COSMOGLOBE DR1 results. III. First full-sky model of polarized synchrotron emission from all WMAP and Planck LFI data

Watts, D. J., et al. 2024c, arXiv:2408.10952, arXiv e-prints, arXiv:2408.10952, Cosmoglobe DR2. I. Global Bayesian analysis of COBE-DIRBE

Willis, J. P., et al. 2020, arXiv:2001.00549, Nature, 577, 39, Spectroscopic confirmation of a mature galaxy cluster at a redshift of 2

Wolz, L., et al. 2017, arXiv:1510.05453, MNRAS, 464, 4938, Erasing the Milky Way: new cleaning technique applied to GBT intensity mapping data

Wuensche, C. A., et al. 2021, arXiv:2106.02107, arXiv e-prints, arXiv:2106.02107, Baryon Acoustic Oscillations from Integrated Neutral Gas Observations: an instrument to observe the 21cm hydrogen line in the redshift range $0.13 < z < 0.45$ – status update

Yang, S., Popping, G., Somerville, R. S., Pullen, A. R., Breysse, P. C., & Maniyar, A. S. 2022, arXiv:2108.07716, ApJ, 929, 140, An Empirical Representation of a Physical Model for the ISM [C II], CO, and [C I] Emission at Redshift $1 \leq z \leq 9$

Yoshiura, S., Shimabukuro, H., Takahashi, K., Momose, R., Nakanishi, H., & Imai, H. 2015, arXiv:1412.5279, MNRAS, 451, 266, Sensitivity for 21 cm bispectrum from Epoch of Reionization

Zaldarriaga, M. & Seljak, U. 1997, astro-ph/9609170, Phys. Rev. D, 55, 1830, All-sky analysis of polarization in the microwave background

Zavala, J. A., et al. 2021, arXiv:2101.04734, ApJ, 909, 165, The Evolution of the IR Luminosity Function and Dust-obscured Star Formation over the Past 13 Billion Years

Zebrowski, J. A. & Stover, A. 2024, SPT-SLIM and VIS: CO Line Intensity Mapping with on-chip mm-wave spectrometers, LIM24 conference, Urbana-Champaign, Illinois, USA. https://mediaspace.illinois.edu/media/t/1_sczrqj3h

Papers

Paper I

COMAP Early Science. III. CO Data Processing

THE ASTROPHYSICAL JOURNAL, 933:184 (21pp), 2022 July 10 https://doi.org/10.3847/1538-4357/ac63ca

COMAP Early Science. III. CO Data Processing

Marie K. Foss[1], Håvard T. Ihle[1], Jowita Borowska[1], Kieran A. Cleary[2], Hans Kristian Eriksen[1], Stuart E. Harper[3], Junhan Kim[2], James W. Lamb[4], Jonas G. S. Lunde[1], Liju Philip[5], Maren Rasmussen[1], Nils-Ole Stutzer[1], Bade D. Uzgil[6,7], Duncan J. Watts[1], Ingunn K. Wehus[1], David P. Woody[4], J. Richard Bond[8], Patrick C. Breysse[9], Morgan Catha[4], Sarah E. Church[10], Dongwoo T. Chung[8,11], Clive Dickinson[3], Delaney A. Dunne[2], Todd Gaier[5], Joshua Ott Gundersen[12], Andrew I. Harris[13], Richard Hobbs[4], Charles R. Lawrence[5], Norman Murray[8], Anthony C. S. Readhead[2], Hamsa Padmanabhan[14], Timothy J. Pearson[2], and Thomas J. Rennie[3]
(COMAP Collaboration)

[1] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway; m.k.foss@astro.uio.no
[2] California Institute of Technology, Pasadena, CA 91125, USA
[3] Jodrell Bank Centre for Astrophysics, Alan Turing Building, Department of Physics and Astronomy, School of Natural Sciences, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[4] Owens Valley Radio Observatory, California Institute of Technology, Big Pine, CA 93513, USA
[5] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[6] California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
[7] National Radio Astronomy Observatory, Pete V. Domenici Array Science Center, P.O. Box 0, Socorro, NM 87801,USA
[8] Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada
[9] Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA
[10] Kavli Institute for Particle Astrophysics and Cosmology & Physics Department, Stanford University, Stanford, CA 94305, USA
[11] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[12] Department of Physics, University of Miami, 1320 Campo Sano Avenue, Coral Gables, FL 33146, USA
[13] Department of Astronomy, University of Maryland, College Park, MD 20742, USA
[14] Departement de Physique Théorique, Université de Genève, 24 Quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
Received 2021 November 19; revised 2022 February 25; accepted 2022 March 15; published 2022 July 13

Abstract

We describe the first-season CO Mapping Array Project (COMAP) analysis pipeline that converts raw detector readouts to calibrated sky maps. This pipeline implements four main steps: gain calibration, filtering, data selection, and mapmaking. Absolute gain calibration relies on a combination of instrumental and astrophysical sources, while relative gain calibration exploits real-time total-power variations. High-efficiency filtering is achieved through spectroscopic common-mode rejection within and across receivers, resulting in nearly uncorrelated white noise within single-frequency channels. Consequently, near-optimal but biased maps are produced by binning the filtered time stream into pixelized maps; the corresponding signal bias transfer function is estimated through simulations. Data selection is performed automatically through a series of goodness-of-fit statistics, including χ^2 and multiscale correlation tests. Applying this pipeline to the first-season COMAP data, we produce a data set with very low levels of correlated noise. We find that one of our two scanning strategies (the Lissajous type) is sensitive to residual instrumental systematics. As a result, we no longer use this type of scan and exclude data taken this way from our Season 1 power spectrum estimates. We perform a careful analysis of our data processing and observing efficiencies and take account of planned improvements to estimate our future performance. Power spectrum results derived from the first-season COMAP maps are presented and discussed in companion papers.

Unified Astronomy Thesaurus concepts: Cosmological evolution (336); CO line emission (262); High-redshift galaxies (734); Molecular gas (1073); Radio astronomy (1338)

1. Introduction

Understanding the evolution of galaxies and the intergalactic medium over the largest spatial and temporal scales is one of the principal goals of cosmology. Galaxy surveys address this challenge by resolving and detecting individual galaxies, a technique that necessarily favors brighter galaxies and smaller cosmic volumes. Spectral line intensity mapping (Madau et al. 1997; Battye et al. 2004; Peterson et al. 2006; Loeb & Wyithe 2008) is a complementary technique (see Kovetz et al. 2017, 2019, for reviews) that holds the potential to characterize the global properties of galaxies and their evolution by surveying the aggregate emission from all galaxies over large volumes.

This technique uses redshifted line emission (e.g., 21 cm, Lyα, CO, or C II) as a tracer for the underlying density field. Large volumes along a given line of sight may be surveyed simultaneously with a single spectrometer at relatively low spatial resolution, and by scanning this spectrometer across the sky, a full 3D density map may be derived. Despite multiple different modeling efforts (Righi et al. 2008; Visbal & Loeb 2010; Lidz et al. 2011; Pullen et al. 2013; Breysse et al. 2014; Li et al. 2016; Padmanabhan 2018; Moradinezhad Dizgah & Keating 2019; Sun et al. 2019; Chung et al. 2022; Moradinezhad Dizgah et al. 2022; Yang et al. 2022) and significant progress on the observational front (Keating et al. 2016, 2020; Riechers et al. 2019; Keenan et al. 2022), the overall level of the CO signal, especially in the clustering regime, is still unknown.

The CO Mapping Array Project (COMAP; Cleary et al. 2022) is an intensity mapping experiment that aims to use emission from carbon monoxide (CO) to trace the aggregate properties of galaxies over cosmic time, back to the Epoch of Reionization. A Pathfinder experiment, consisting of a 19-feed 26–34 GHz receiver, has been fielded on a 10.4 m single-dish telescope at the Owens Valley Radio Observatory (OVRO).[15] In this frequency range, the receiver is sensitive to CO (1–0) at $z = 2.4$–3.4, with a fainter contribution from CO (2–1) at $z = 6$–8. The main goal of the Pathfinder is to detect the CO (1–0) signal and use it to constrain the properties of galaxies at the Epoch of Galaxy Assembly. A future phase will add a second receiver at 12–20 GHz in order to detect CO (1–0) from around $z = 5$–9, cross-correlating with the CO (2–1) signal from the 26–34 GHz receiver and constraining the properties of galaxies toward the end of the Epoch of Reionization.

The receiver's detector chain is based on cryogenically cooled HEMT low-noise amplifiers (LNAs) that contribute to a typical system temperature of about 44 K across the full frequency range (see the Appendix for more details). The predicted signal from high-redshift CO emission is expected to be no more than a few microkelvin per COMAP spatial/spectral resolution element (or "voxel"). Thus, the noise must be reduced by many orders of magnitude, compared to the raw instrumental noise, before a statistically significant detection may be achieved. In practice, this is done by repeatedly observing the same part of the sky using multiple detectors, and thereby gradually increasing the sensitivity per voxel. For this to succeed, however, it is necessary to suppress systematic contributions from atmospheric temperature variations, sidelobe contamination, ground pickup, standing waves, Galactic foregrounds, etc., by a corresponding amount.

The first-season COMAP science observations started in 2019 June and lasted until 2020 August. This paper describes the first-season COMAP data analysis pipeline, which aims to produce clean maps from raw time-ordered COMAP observations. This includes calibration, data selection, filtering, and mapmaking. The rest of this paper is organized as follows: First, in order to establish useful notation and conventions, we give a brief introduction to the COMAP instrument in Section 2, while referring the interested reader to Lamb et al. (2022) for full details. Next, we provide a high-level overview of the analysis pipeline in Section 3.1, before specifying each step in Sections 3.3–3.6. Data selection and efficiency are discussed in Sections 4 and 5. The results are presented in Section 6, and we summarize and conclude in Section 7.

2. Instrument and Data Model

Before describing the COMAP analysis pipeline, we provide a brief overview of the instrument itself and define an explicit data model. A more detailed description of the instrument can be found in a separate paper (Lamb et al. 2022).

[15] https://www.ovro.caltech.edu/

2.1. Instrument Overview

The COMAP Phase I instrument observes in the K*a* band, at 26–34 GHz, and is located at the OVRO in California, USA. It is mounted on a 10.4 m telescope that was originally built for the Millimeter Array at OVRO and then used as a part of the Combined Array for Research in Millimeter-wave Astronomy (Woody et al. 2004) experiment, and it has now been repurposed for COMAP. The telescope's primary and secondary reflectors have diameters of 10.4 and 1.1 m, respectively, and the beam FWHM is about 4$'$.5 at 30 GHz.

The receiver comprises 19 independent detector chains, called "feeds." The signal chain of each feed consists of individual feed horns, polarizers, LNAs, two stages of downconversion, frequency separation, and digitization. For the observations described in this paper, 15 feeds have a two-stage polarizer, two feeds have a single-stage polarizer, and two feeds have no polarizer. The digitization happens in two CASPER "ROACH-2" FPGA-based spectrometers for each signal chain, giving us four 2 GHz wide sidebands, each of which has 1024 frequency channels, resulting in a native frequency resolution of approximately 2 MHz. The two sidebands of each band (A and B) are labeled "lower" (LSB) or "upper" (USB). For more details on the instrument see Lamb et al. (2022).

To support frequent and accurate gain estimation, COMAP employs an ambient temperature load that is directly attached to the environmental shroud housing. This "calibration vane" is automatically moved in front of the feed-horn array at the beginning and end of each observation (each lasting for about 1 hr; see Section 2.3), fully filling the field of view of each pixel. The temperature of the calibration vane is monitored with sensors, allowing the system temperature to be calculated and applied to calibrate the gain (see Section 3.4 for more details).

2.2. Field Selection

COMAP observes several parts of the sky. Table 1 lists all CO science fields and calibrators. In Figure 1 we plot the elevation of the CO and calibration fields as a function of Local Sidereal Time, indicating when the fields are available for observation. The three CO fields were selected to maximize the observing efficiency, avoid bright 30 GHz point sources ($\gtrsim$1 Jy), and overlap with the coverage of the Hobby–Eberly Telescope Dark Energy eXperiment (HETDEX; Hill et al. 2008, 2021; Gebhardt et al. 2021), a galaxy survey targeting Lyα emission from galaxies in the same redshift. Although COMAP's observing strategy has been designed to permit the direct detection of CO fluctuations from galaxies at $z = 2.4$–3.4, cross-correlation with a galaxy survey such as HETDEX can increase the detection significance by at least a factor of two (Chung et al. 2019; Silva et al. 2021), as well as provide validation for the origin of detected signal in galaxies at the target redshift.

In addition to the main science fields, we are also conducting a survey of the Galactic plane covering longitudes $20° < l < 220°$, details of which can be found in Rennie et al. (2022).

To facilitate calibration with astrophysical sources, we observe a handful of radio sources, including Jupiter, the supernova remnants Taurus A (TauA) and Cassiopeia A (CasA), and the radio galaxy Cygnus A (CygA), all of which are somewhat extended compared to the beam, except for

Table 1
COMAP Fields and Calibrators

Field Name	R.A. (J2000)	Decl. (J2000)	Notes
Field 1	01:41:44.4	+00:00:00.0	CO science field—lies within the HETDEX Fall field
Field 2	11:20:00.0	+52:30:00.0	CO science field—lies within the HETDEX Spring field
Field 3	15:04:00.0	+55:00:00.0	CO science field
TauA	05:34:31.9	+22:00:52.2	Pointing calibrator—supernova remnant (Crab Nebula)
CasA	23:23:24.0	+58:48:54.0	Pointing calibrator—supernova remnant
CygA	19:59:28.4	+40:44:02.1	Pointing calibrator—radio galaxy
Jupiter			Pointing calibrator

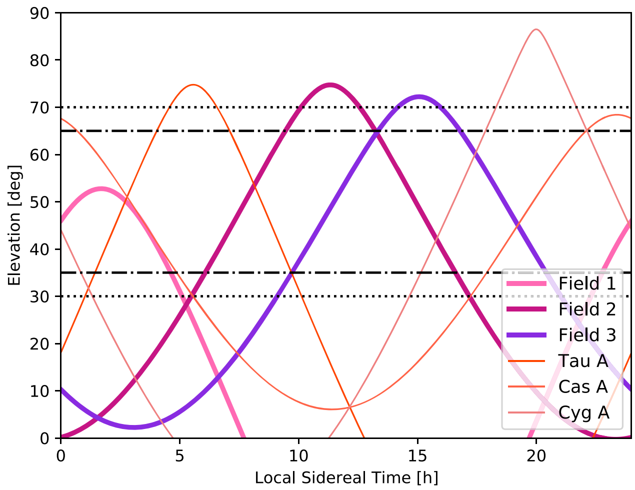


Figure 1. Elevation of CO (pink/purple) and calibration (orange) fields as a function of local sidereal time.

Jupiter. These astrophysical calibrators are used to determine the overall normalization of the beam transfer function. See Ihle et al. (2022) and Rennie et al. (2022) for more details.

2.3. Observation Strategy

Telescope scans of the science fields follow a harmonic motion described by

$$\mathrm{az} = A\sin(at + \phi) + \mathrm{az}_0; \quad \mathrm{el} = B\sin(bt) + \mathrm{el}_0, \tag{1}$$

where A, B are amplitude parameters that determine the angular extent of the scan, the ratio a/b determines the shape of the curve, and ϕ is a phase parameter. az_0 and el_0 correspond to the sky position of the field center midway through the scan. Two different scan types were used. "Lissajous" scans are performed with the following parameters:

$$A = 48'/\cos(\mathrm{el}_0), \tag{2}$$

$$B = 48', \tag{3}$$

$$a = 39'\,\mathrm{s}^{-1}/A, \tag{4}$$

$$b = 19\rlap{.}'5\,\mathrm{s}^{-1}/B, \tag{5}$$

$$\phi = \pi/2. \tag{6}$$

Note the factor of $1/\cos(\mathrm{el}_0)$, which makes sure the scan area is roughly the same at all elevations. Note also that the time parameter, t, is set to zero at the start of each scan. "Constant elevation scans" (CES) use the same parameters in azimuth but have no movement in elevation (i.e., $B = b = 0$). At the start of a scan, the telescope is positioned at the leading edge of the field. The telescope then executes the scan while the field drifts through the pattern. This typically takes 3–10 minutes, after which the telescope is repointed to the leading edge of the field again in preparation for the next scan. An example of the scanning path for about 1 hr of continuous observations with a Lissajous scan and a CES is shown in Figure 2. Testing the relative performance of the CES and Lissajous scanning strategies in terms of final data quality is an important goal of the first-season COMAP survey.

2.4. Data Model

As described by Lamb et al. (2022), the COMAP detector readout for a single-frequency channel may be modeled as

$$P_{\mathrm{out}} = k_{\mathrm{B}} G \Delta\nu T_{\mathrm{sys}}, \tag{7}$$

where k_{B} is the Boltzmann constant, G is the gain, $\Delta\nu$ is the bandwidth, and T_{sys} is the system temperature of the instrument. The system temperature may be further modeled as[16]

$$T_{\mathrm{sys}} = T_{\mathrm{receiver}} + T_{\mathrm{atmosphere}} + T_{\mathrm{ground}} + T_{\mathrm{CMB}} + T_{\mathrm{foregrounds}} + T_{\mathrm{CO}}, \tag{8}$$

where T_{reciever} is the effective noise temperature of the receiver, $T_{\mathrm{atmosphere}}$ is the noise contribution from the atmosphere, T_{ground} is ground pickup, T_{CMB} is the contribution from the cosmic microwave background (CMB), $T_{\mathrm{foregrounds}}$ are continuum foregrounds (typically from the galaxy), and T_{CO} is the line emission signal from extragalactic CO, which is the main scientific target of the COMAP instrument.

To understand the challenges involved in measuring the cosmological CO signal, it is instructive to consider the order of magnitude and stability of each term in Equation (8). The largest single contribution is that of the receiver temperature, which is usually about 10–30 K. For the COMAP receiver, with HEMT LNA technology, this is very stable.

The second-largest contribution is from the atmosphere, which typically adds 15–25 K. This term varies significantly on all timescales longer than a few seconds and depends on external conditions, including elevation, humidity, cloud coverage, ambient temperature, and wind speed. It is also

[16] In this section we are writing all the contributions to T_{sys} in terms of their effective noise contribution, rather than any physical temperatures. See Section 3.4 for a definition of T_{sys} in terms of physical quantities.

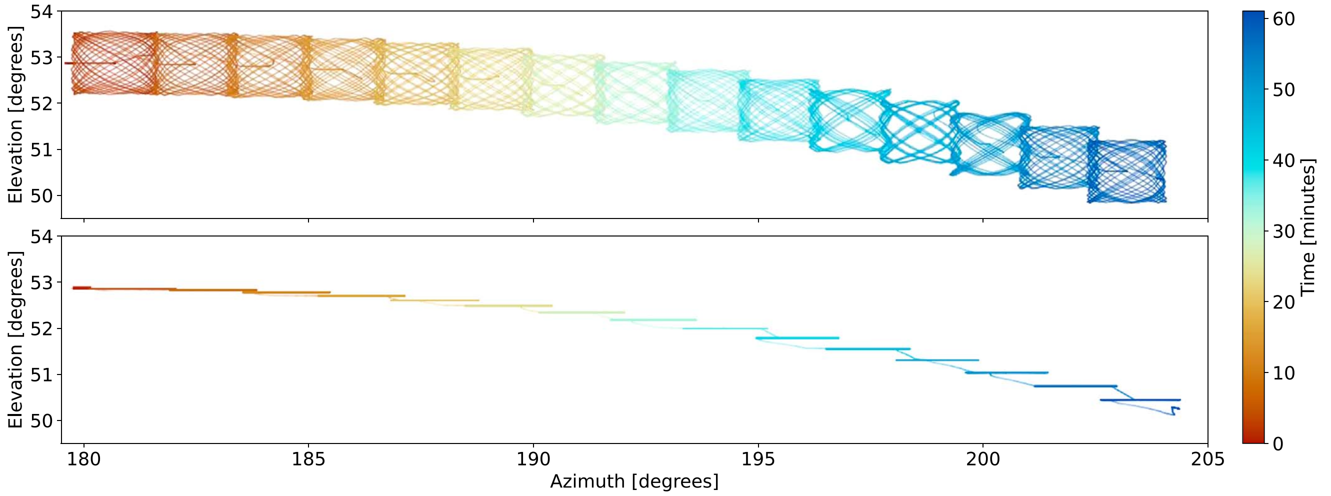


Figure 2. Movement of the telescope boresight in azimuth and elevation for an observation employing Lissajous scans (top) and an observation employing CES (bottom). Both observations consist of 15 individual scans of Field 1.

strongly correlated between detectors and frequencies, since all feeds observe through essentially the same atmospheric column at any given time; fortunately, the phase structures of the atmospheric fluctuations are uncorrelated on long timescales.

Next, ground pickup typically accounts for 5–6 K. Most of this contribution is from illumination spillover around the primary, i.e., ground signal diffracting at the edge of the secondary past the edge of the primary. A secondary contribution from ground signal reflecting off the secondary support legs, however, can be particularly problematic because it depends sensitively on the instrument pointing: if a sidelobe happens to straddle a strong signal gradient, such as the horizon or the Sun, several millikelvin variations may be measured on very short timescales and with a time dependency that appears nearly sky synchronous.

The fourth term represents the CMB temperature of 2.7 K, which is both isotropic and stationary, while the fifth term represents astrophysical foregrounds, expected to contribute at most 1 mK, for instance, synchrotron, free–free, and dust emission from the Galaxy. Although these are sky synchronous and, in principle, could confuse potential CO measurements, they also have very smooth frequency spectra (Keating et al. 2015) and are therefore relatively easy to distinguish from the cosmological CO signal, which varies rapidly with frequency. An important potential exception is line emission from other molecules redshifted to our band from galaxies at other epochs. The hydrogen cyanide (HCN) line is expected to be one of the brightest such lines. Emission from HCN in galaxies toward our CO fields at redshift z = 1.6–2.4 will appear in our frequency range. However, this contribution is expected to be an order of magnitude lower than that from CO (Chung et al. 2017).

Finally, the cosmological CO line emission signal is expected to account for $\mathcal{O}(1\ \mu\mathrm{K})$. Whether it is possible to detect such a weak signal depends directly on the stability and sensitivity of the instrument. In this respect, the fundamental quantity of interest is the overall noise level of the experiment, which is dominated by random thermal noise.

The magnitude of these random thermal fluctuations is proportional to T_{sys}, with a standard deviation that is given by the so-called radiometer equation,

$$\sigma_{\mathrm{N}} = \frac{T_{\mathrm{sys}}}{\sqrt{\Delta\nu\ \tau}}, \tag{9}$$

where τ is the integration time. Thus, since both the system temperature and the bandwidth are essentially fixed experimental parameters, the only way of reducing the total uncertainty is by increasing the integration time. As a concrete and relevant example, we note that an integration time of 45 hr is required to achieve a standard deviation of 20 μK with a system temperature of 45 K and a bandwidth of 31.25 MHz.

In addition to the thermal and uncorrelated noise described by the radiometer equation, there are three main sources of correlated noise, namely, gain fluctuations in the LNAs, atmospheric temperature fluctuations, and time-dependent standing waves. All of these are expected to have a roughly $1/f$-type spectrum, although with different particular properties.[17] The fact that these sources of correlated noise are also strongly correlated between frequencies is very useful in order to filter out this noise in the analysis.

Equation (7) describes the detector output at any given time. To connect this to the actual measurements recorded by the detector, we adopt the following data model:

$$\begin{aligned} d_\nu^i(t) = \langle d_\nu^i\rangle(1 + \delta_G^i(t))[1 + P_{\mathrm{cel}}^i(\Delta s_{\mathrm{cont}} + \Delta s_{\mathrm{CO}}^\nu) \\ + P_{\mathrm{tel}}^i \Delta s_{\mathrm{ground}} + n_{\mathrm{corr}}(t) + n_w^{\nu i}(t)]. \end{aligned} \tag{10}$$

Here $d_\nu^i(t)$ denotes the raw data recorded at time t for frequency channel ν in feed i; $\langle d_\nu^i\rangle$ represents the corresponding time average and basically corresponds to $\langle T_{\mathrm{sys}}^{i\nu}(t)\rangle\langle G_\nu^i(t)\rangle$; $\delta_G^i(t)$ denotes feed-dependent gain fluctuations; P_{cel}^i and P_{tel}^i are pointing matrices in celestial and telescope coordinate systems, respectively; Δs_{cont} denotes the celestial continuum source fluctuations, mainly from the CMB and Galactic foregrounds; $\Delta s_{\mathrm{CO}}^\nu$ is the CO line emission fluctuation; $\Delta s_{\mathrm{ground}}$ is the ground signal fluctuation picked up by the far sidelobes; and $n_{\mathrm{corr}}(t)$ are the correlated temperature fluctuations, mostly

[17] There are several different sources of standing waves; some of the main ones give rise to $1/f$-like spectra, but others do not.

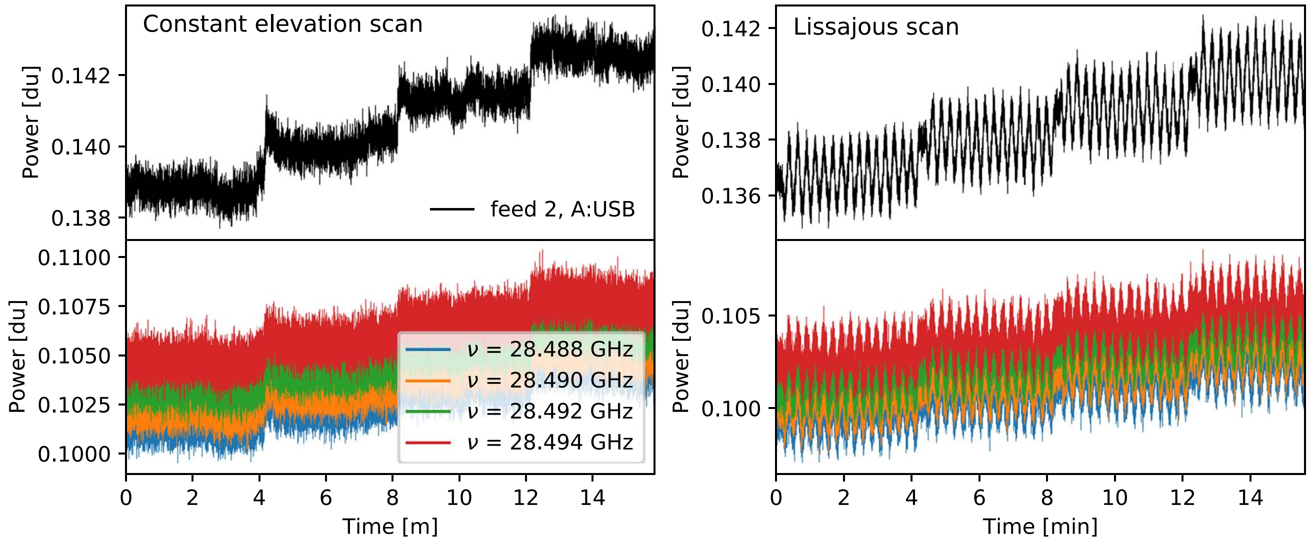


Figure 3. Raw data from the COMAP instrument (in arbitrary digital units of power). Here we see data averaged over a single 2 GHz wide sideband (top) and examples of data from four individual frequency channels in that sideband (bottom). These data were taken using two different scan patterns: CES (left) and Lissajous (right).

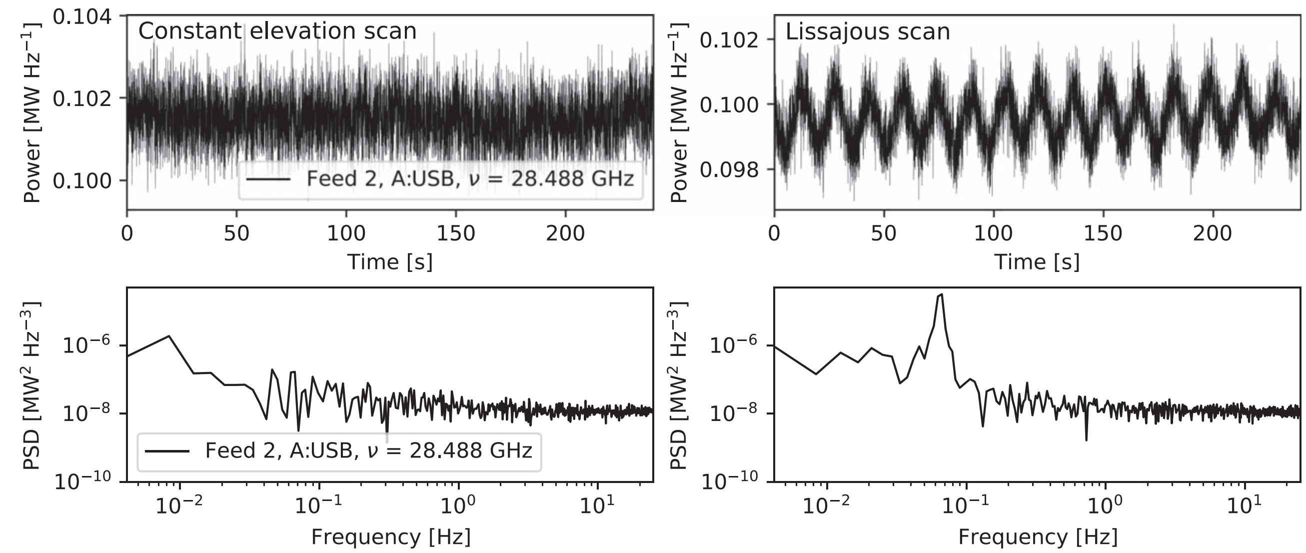


Figure 4. Raw data from an individual frequency channel of the COMAP instrument. Power is shown as a function of time (top), and the corresponding PSD is also shown (bottom). We show data from a CES (left) and a Lissajous scan (right).

consisting of atmosphere fluctuations and standing waves. Factors with no feed or frequency index are assumed to be similar (or at least strongly correlated) at different frequencies and feeds, while factors with a ν label indicate parts of the model that are assumed to have nonsmooth frequency dependence. The main purpose of the COMAP analysis pipeline is to characterize $\Delta s_{\mathrm{CO}}^{\nu}$ given $d_{\nu}^{i}(t)$.

2.5. Data Overview

Before presenting the analysis pipeline, we provide a preview of the raw time-ordered data (TOD) generated by the COMAP instrument, with the goal of building intuition that will be useful for understanding the purpose of each component of the analysis pipeline described in this paper. Figures 3 and 4 show examples of such raw TOD from the instrument using the CES (left column) and Lissajous (right column) scanning strategies. Perhaps the most obvious features in these plots are stepwise changes in power as the telescope changes elevation during repointings between scans; see Section 2.3. The Lissajous scans additionally show oscillations in power as the telescope changes elevation during the scan, since the telescope looks through a thicker slab of atmosphere at lower elevations, and this increases the atmospheric contribution to the system temperature.

The top panels in Figure 4 show an individual frequency channel for a single scan (i.e., stationary observation period), while the bottom panel shows the corresponding power spectral density (PSD). For the CES case, the PSD is relatively featureless, with an overall shape that looks consistent with a typical $1/f$ noise spectrum. For the Lissajous case, an additional strong peak is seen around 0.007 Hz, which matches the scanning period of 14 s, and this corresponds to the periodic atmospheric variations seen in the panels above.

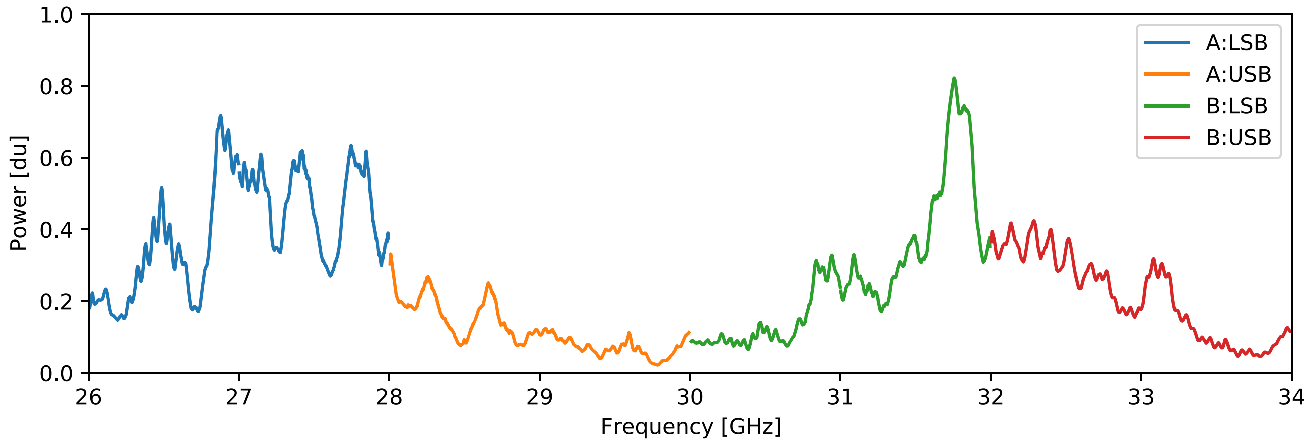


Figure 5. Time-averaged raw data from each frequency channel on a single feed of the COMAP instrument. The colors represent the four 2 GHz wide sidebands. Note that a few of the frequency channels at the edges and middle of sidebands tend to be unstable and are masked out in the analysis.

Figure 5 shows the time-averaged data for all frequency channels of a single feed for one scan. The spectral shape is mostly determined by the average gain as a function of frequency, due to the combined effect of the various components of the receiver chain. This average gain is a purely instrumental effect, not associated with the true sky signal, and therefore simply corresponds to a normalization factor that should be calibrated out before higher-level analysis. However, some of the spectral shape is also determined by the fact that the system temperature also changes with frequency, and in some cases exhibits large spikes within specific frequency ranges (see Lamb et al. 2022 for more details). Separating the gain variation as a function of frequency from the system temperature as a function of frequency is a main goal of the calibration procedures described in Section 3.4.

In Figure 6 we plot the correlation,

$$C_{ij} = \frac{\langle \hat{d}^i \hat{d}^j \rangle}{\sqrt{\langle \hat{d}^i \hat{d}^i \rangle \langle \hat{d}^j \hat{d}^j \rangle}}, \tag{11}$$

between the power, $\hat{d}^i$, recorded by any two feeds, i and j, after averaging over all frequencies within each sideband for each radiometer. Here we first note that the data from different sidebands of the same feed are strongly correlated. This is because both main sources of correlated noise in the COMAP data, namely, gain fluctuations and atmospheric fluctuations, are common for sidebands within a given feed. In contrast, sidebands for different feeds mostly share the atmospheric fluctuations (and also some standing waves) but have independent gain fluctuations, and this results in lower overall correlations, but still typically in the 10%–40% range. Accounting for and mitigating such correlations will clearly be essential in order to extract robust science from these observations.

The quality of the COMAP data depends strongly on the observing conditions, as illustrated in Figure 7. The top panel shows an observation made under normal conditions, while the middle panel shows an observation made during poor weather, with thick cloud coverage. The bottom panel shows a data segment with strong "spikes," a feature of some data taken during summer. Such spikes have been seen to occur when insects are flying in front of the receiver. Automatic identification and removal of problematic data is clearly an important and necessary component of the pipeline.

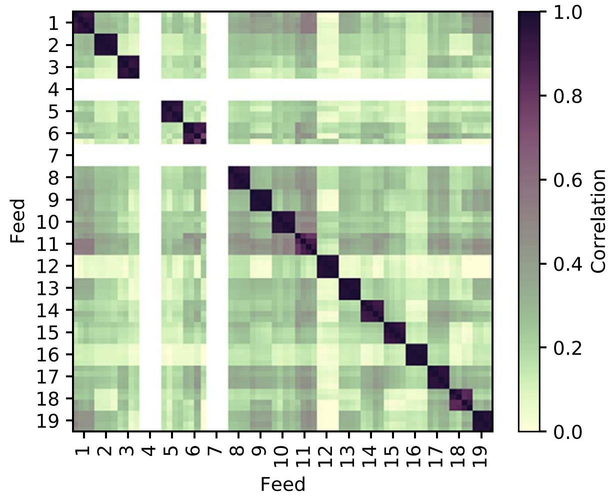


Figure 6. Correlation between the sideband-averaged data from the 19 feeds of the COMAP instrument for a single CES. Note that within each feed-feed cell there are subcells showing the correlations between individual sidebands. For this observation, as for much of the observing campaign, the LNAs for feeds 4 and 7 were turned off because those feeds, as a test, did not have a polarizer and so had large standing waves owing to reflections between the receiver and the secondary reflector.

Finally, Figure 8 shows the calibration vane observations that are made at the beginning and end of each observation period. Since the ambient temperature is about one order of magnitude higher than T_{sys}, the measured power is also correspondingly about one order of magnitude higher, and this bright and known signal allows for a precise estimate of T_{sys}. Note that these data segments are removed prior to data analysis, as they would otherwise compromise any filtering that may be applied to the data.

3. COMAP Analysis Pipeline

3.1. Pipeline Overview

We are now ready to present the COMAP analysis pipeline, which is designed to process the raw data discussed in Section 2.5 into calibrated and cleaned CO maps. The main

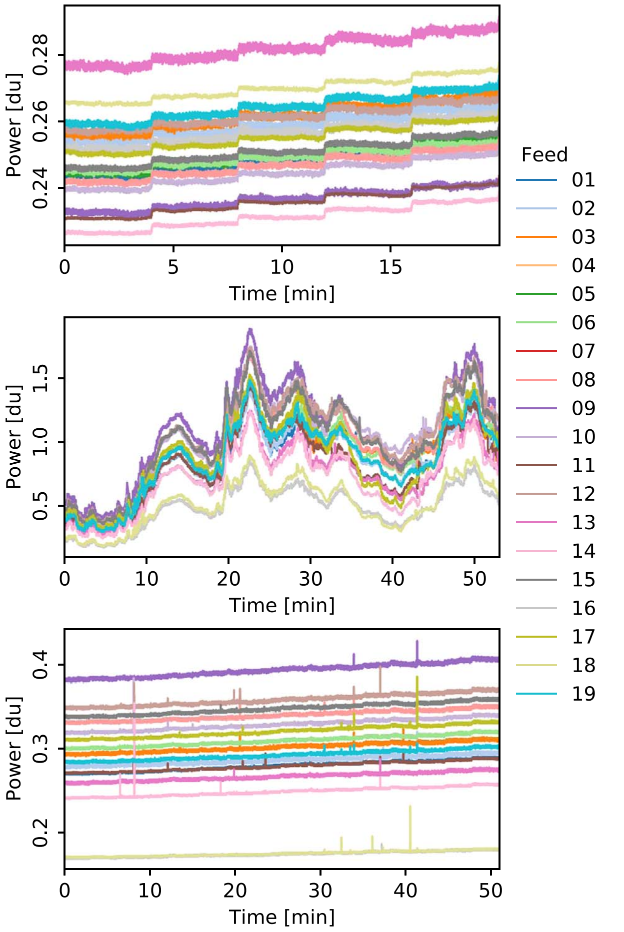


Figure 7. Feed-averaged COMAP TOD recorded under various observing conditions. The top panel shows data observed under normal conditions and is dominated by instrumental noise. The middle panel shows data observed under poor weather conditions with a thick cloud coverage, resulting in large coherent power fluctuations observed by all feeds. The bottom panel shows data with strong spikes, which may, for instance, happen during rare periods with high insect activity.

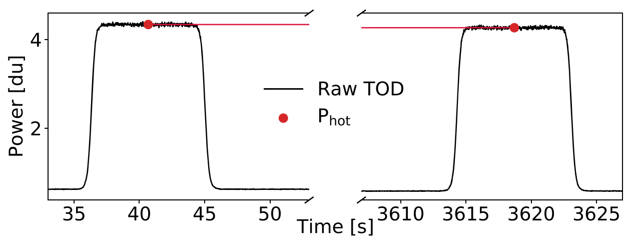


Figure 8. The calibration vane is inserted in front of the receiver at the beginning and end of one observation of a CO science field. The time between calibration vane insertions is typically about an hour, a period set by the preferred data file size for the CO field observations.

steps of this pipeline are schematically illustrated in Figure 9, and the corresponding codes are listed in Table 2.

The processing starts with "Level 1" files, which contain raw data as recorded by the instrument, together with pointing information and housekeeping data. Each of these files typically contains about 1 hr of observation time, including calibration vane observations at the beginning and end. We denote each (roughly) 1 hr of data as one observation and assign it an individual observation ID (abbreviated obsID). Each observation consists of several scans, where one scan is the period between two repointings of the telescope, during which the telescope performs the same motions around a fixed point in azimuth and elevation while the target field drifts through. The instrumental properties are consequently assumed to be stationary within each scan. The module denoted `scan_detect` in Figure 9 indicates a dedicated code that partitions each observation into individual scans, based on pointing information, and records information of each scan in a database.

The main processing takes place in the `l2gen` module, which generates calibrated and cleaned TOD and stores them in so-called "Level 2" files. This is achieved through the application of a series of filters (see Section 3.3) and a time-varying gain normalization (see Section 3.4). This stage also evaluates basic goodness-of-fit statistics and defines a frequency channel mask that excludes missing or broken data for the current scan, before reducing the spectral resolution of the data to a spectral resolution suitable for mapmaking. This demonstrates the advantage of the high spectral resolution of the raw data. While our cosmological signal does not have much structure on scales corresponding to these high resolutions, the systematic effects do. The high resolution thus allows us to filter out or mask systematic effects more precisely, without masking entire low-resolution frequency channels. In our main analysis, we reduce the resolution from $\sim$2 to $\sim$31 MHz, resulting in the computational speedup of subsequent steps and a memory saving for storing final maps by a factor of 16.

Next, the `accept_mod` module reads in the statistics (including goodness of fit) and basic frequency mask produced by `l2gen` and produces a list of accepted observations as defined by user-specified thresholds for each statistic (see Section 4). Examples of relevant statistics used for this purpose are χ^2 per observation, correlated noise knee-frequency ($f_{\rm knee}$), and solar elongation. The output from this process is called an *accept list*, which determines what data to use for mapmaking.

Converting TOD into pixel-ordered data is done by a mapmaker called `tod2comap` (see Section 3.6). As shown in the following sections, the adopted filters result in very nearly uncorrelated white noise, and the current implementation of `tod2comap` accordingly adopts simple binning into voxels. Finally, from these maps we can estimate the CO power spectrum using the module `comap2ps` (see Ihle et al. 2022 for details).

3.2. Data Segmentation

As described above, we define a *scan* to be the observing period between repointings of the telescope. The purpose of the `scan_detect` code is to identify all scans within all observation periods and produce an observation database, consisting of a list of obsIDs sorted according to source. For each obsID, we list all scans within that obsID, including basic information such as the Modified Julian Date (MJD) of the start and end of the scan, as well as the scanning mode (e.g., Lissajous or CES) and mean pointing information.

3.3. Filtering

As described in Section 3.1, the COMAP TOD exhibit a wide range of non-CO-related contributions, both of instrumental and external origin. These must be suppressed by orders

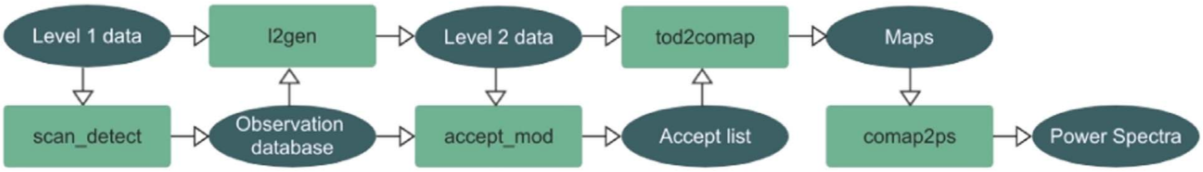


Figure 9. Flow diagram of the analysis pipeline. The dark-green ellipses are data products, whereas the light-green boxes are the different modules of the data analysis pipeline.

Table 2
Analysis Pipeline Software Routines

Module Name	Input	Output	Description
`scan_detect`	Level 1 files	Obs. database	Classifies and gathers info for the required set of observations
`l2gen`	Level 1 files, Obs. database	Level 2 files	Performs filtering and calibration of the TOD
`accept_mod`	Level 2 files, Obs. database	Accept list	Performs data selection
`tod2comap`	Level 2 data, Accept list	Maps	Converts TOD into 3D maps
`comap2ps`	Maps	Power spectra	Calculates and combines auto- or cross-spectra from maps

of magnitude prior to mapmaking in order to extract the astrophysically valuable signal. With this goal in mind, we introduce four specific filters, each targeting one class of artifacts.

Figure 10 shows the evolution of the data as it passes through each of the filters.

3.3.1. Normalization

The first filtering operation we introduce is data normalization. This is done simply by dividing the raw TOD, $P_{\rm out}$, by its own running mean and then subtracting 1,

$$d(\nu, t) = \frac{P_{\rm out}(\nu, t)}{\langle P_{\rm out}(\nu, t)\rangle} - 1. \quad (12)$$

Here t is a time sample index and ν denotes frequency channel. This operation is performed separately on each frequency channel. The running mean is estimated by putting the data through a low-pass filter with a timescale of about 100 s. This step basically removes $\langle d_\nu^i\rangle$ from Equation (10), and it also removes the first term in square brackets (which is equal to 1) of the same equation.

The main purpose of this step is to equalize (i.e., "flatten") the instrumental passband, as illustrated in Figure 5, and effectively establish data with appropriate relative calibration. The main practical advantage of doing so is that the amplitude of common-mode contaminants, such as gain-induced correlated noise or atmospheric fluctuations, becomes comparable across all frequencies within a single sideband, and therefore much easier to filter out. The same also holds true for broadband astrophysical contributions, such as the CMB or foregrounds, which also must be removed prior to signal extraction. See the top panel of Figure 10 to see the effect of the normalization step. We can see that long-timescale fluctuations are removed and that the data now fluctuate around zero.

Note also that with the definition in Equation (12) the noise level of $d(\nu, t)$ is given by the sample rate and bandwidth alone in the ideal case and should equal $1/\sqrt{\tau\Delta\nu}$. Calibration into physical units is performed simply by multiplying $d(\nu, t)$ by $T_{\rm sys}$. We find that $d(\nu, t)$ is a particularly convenient function for goodness-of-fit tests, and it will serve as our main object of interest in the following.

3.3.2. Removal of Az/El Templates

The second filter we apply is designed to suppress signals that are correlated with local pointing (azimuth and elevation), as opposed to sky-correlated signals. The two main effects of this type are elevation-correlated atmospheric contributions and azimuth-correlated sidelobe contributions. The first of these effects may be modeled by a simple expression for the optical depth of the atmosphere of the form

$$\tau(\mathrm{el}) = \frac{\tau_0}{\sin(\mathrm{el})}, \quad (13)$$

where τ_0 is the optical depth of the atmosphere at zenith and el is the elevation, while the second effect may be approximated through a low-order polynomial in azimuth. We therefore filter the data by fitting and subtracting the following simple model to each normalized frequency channel separately:

$$d = \frac{g}{\sin(\mathrm{el}(t))} + a\,\mathrm{az}(t) + c + n. \quad (14)$$

Here g, a, and c are fitting constants, and n denotes Gaussian noise with an assumed constant variance. We find the best-fit values for the free parameters by minimizing a χ^2 statistic, and we use g and a to clean the TOD with respect to the Az/El templates,

$$d_{\rm after} = d_{\rm before} - \frac{g}{\sin(\mathrm{el})} - a\,\mathrm{az} - \left\langle \frac{g}{\sin(\mathrm{el})} + a\,\mathrm{az}\right\rangle. \quad (15)$$

In this expression, $\langle\ \rangle$ denotes the mean value in time for a specific frequency channel, and this term ensures that the TOD has vanishing mean also after subtraction of Az/El templates. For long-duration scans we divide the TOD into disjoint segments of roughly 4 minutes each and perform the template fit and removal separately on each data segment, in order to improve the tracking of temporal variations.

The effect of the Az/El template removal can be seen in the second row of Figure 10.

3.3.3. Polynomial Continuum Filter

Our third filter, usually referred to as the "polyfilter," is designed to remove all continuum signals observed

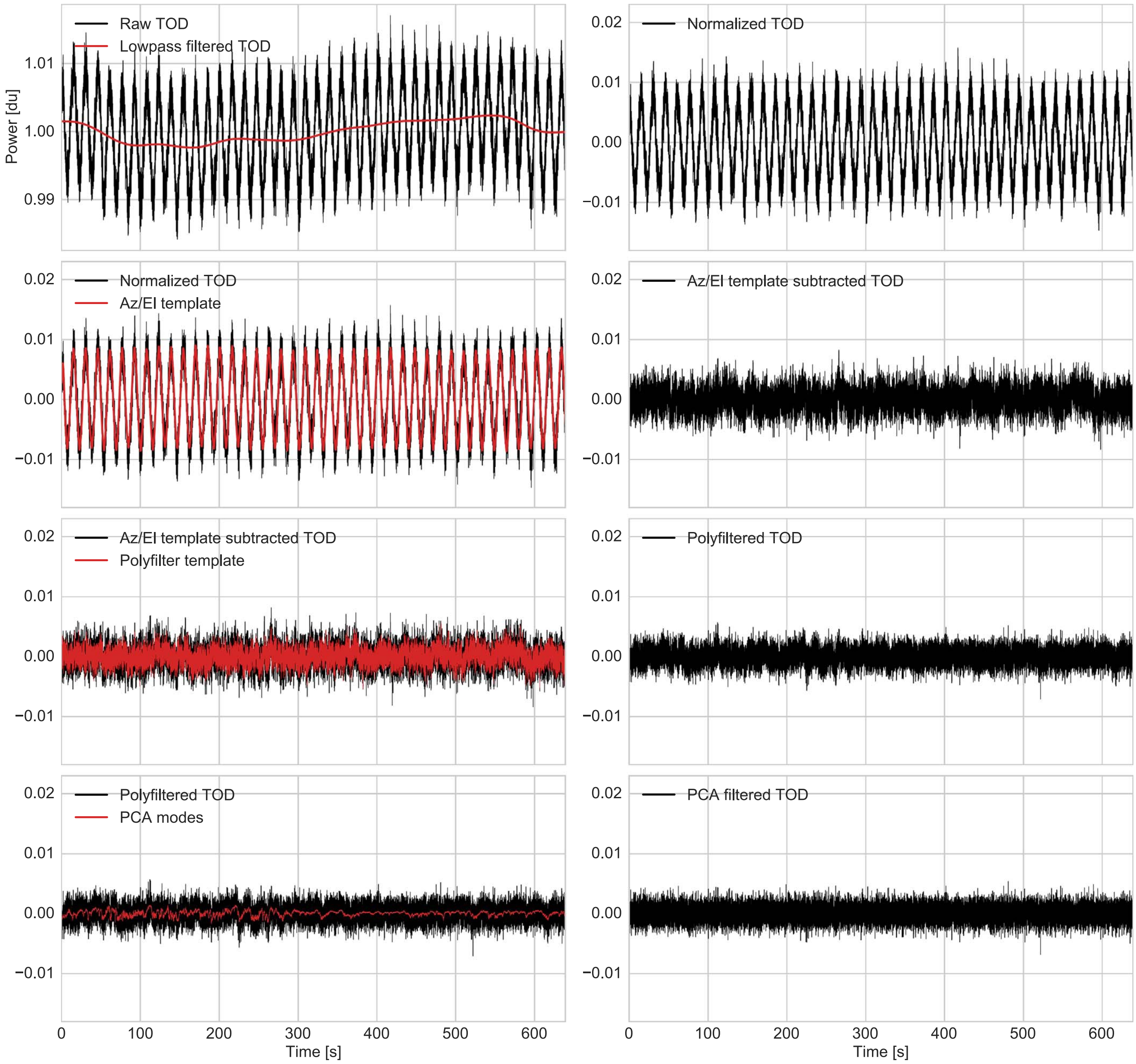


Figure 10. Effect of each filter in time domain. Each row shows the data before (left column) and after (right column) applying the indicated filter. From top to bottom, the filters shown are (1) normalization, (2) elevation gain subtraction, (3) polyfiltering, and (4) PCA filtering. Data used are from scan 7717.03, feed 9, in a 31.25 MHz band around 32.3 GHz.

simultaneously by all frequency channels within a given sideband. Specifically, for each time step we fit and subtract a low-order (and typically linear) polynomial to the normalized and Az/El-subtracted TOD in frequency space for each sideband.

We assume

$$d_\nu = c_0 + c_1\nu + c_2\nu^2 + ..., \tag{16}$$

where d_ν are the data across one sideband at a specific time step and c_0, c_1, c_2, etc., are constants that are fitted independently for each sideband and at each time step. We then remove the fitted polynomial from the data. In the third row of Figure 10, we can see an example of how this filter removes the majority of the correlated noise from the data.

The main target of this filter is $1/f$ noise from gain variations in the receiver electronics and atmospheric temperature fluctuations, which is strongly correlated between frequency channels within each sideband. Indeed, the fact that this noise is so tightly correlated between channels is one of the key instrumental features of the COMAP instrument that makes CO measurements feasible in the first place, effectively reducing the final noise level by a significant amount.

As a bonus, this polynomial filter also suppresses any slowly varying astrophysical signal, and in particular broadband signals such as CMB, synchrotron, free–free, or anomalous microwave emission. In contrast, the cosmological CO signal is expected to vary on the scale of adjacent frequency channels and is therefore only mildly affected by this filter. However, some CO signal is indeed lost on the largest longitudinal scales

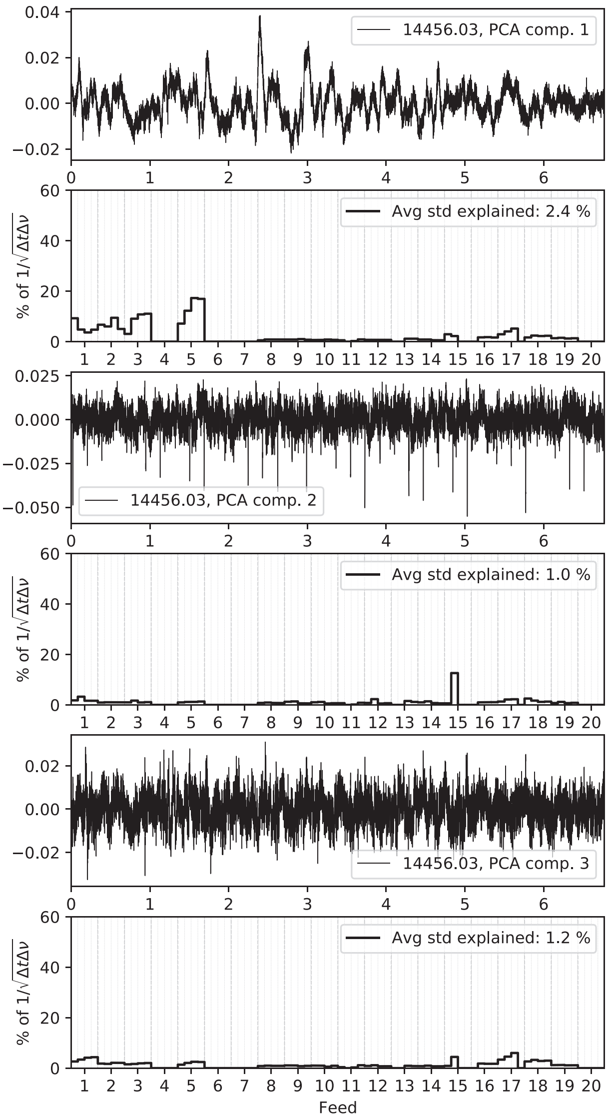


Figure 11. The three leading PCA components of a typical scan, and which feeds are affected.

as a result of this filter, and this effect will later be quantified in terms of an effective transfer function (see Section 5 for more details).

3.3.4. Principal Component Analysis Filter

While the previous filter removes continuum signals within each sideband, our fourth and final filter targets common-mode signals seen simultaneously by the entire focal plane. The two most prominent examples of such contaminants are residual atmospheric variations and standing waves, both of which have strongly correlated time variations across all feeds and frequencies. To suppress these signals, we perform a so-called principal component analysis (PCA) on the whole data set and subtract the leading modes. Intuitively speaking, this amounts to identifying the functions of time that explain the largest amount of the variance between the different frequencies across all the different feeds. These functions are the leading PCA components.

To formulate this idea in a mathematical language, let us organize all data in a given scan into a data matrix D, where each row contains the TOD corresponding to a single-frequency channel on a single feed. Thus, D is a matrix with dimensions $n_{\rm freq} \times n_{\rm samp}$, where $n_{\rm freq} = n_{\rm feeds} \cdot n_{\rm sidebands} \cdot n_{\rm freqpersideband} = 19 \cdot 4 \cdot 1024$ is the total number of frequency channels added up from all sidebands and feeds and $n_{\rm samp}$ is the number of samples in time, such that

$$D = \begin{bmatrix} D_{11} & \cdots & D_{1n_{\rm samp}} \\ \vdots & \ddots & \vdots \\ D_{n_{\rm freq}1} & \cdots & D_{n_{\rm freq}n_{\rm samp}} \end{bmatrix}. \tag{17}$$

The empirical data covariance matrix, C, may then be written as

$$C = D^T D, \tag{18}$$

and the eigenvectors, $\boldsymbol{v}_k$, of this matrix that correspond to the highest eigenvalues are precisely the PCA components we are looking for. In practice, we identify the few leading PCA components through a standard iterative method.

For each frequency (in each feed) we compute the PCA amplitudes by projecting the observed data vector, $\boldsymbol{d}$, onto the PCA eigenvector,

$$a_k = \boldsymbol{d} \cdot \boldsymbol{v}_k = \sum_{i=1}^{n_{\rm samp}} d_i v_k^i, \tag{19}$$

where $\boldsymbol{d}$ is now the normalized, Az/El-template-subtracted, and polynomial-filtered data described above. The leading PCA components are then subtracted from the data,

$$\boldsymbol{d}_{\rm after} = \boldsymbol{d}_{\rm before} - \sum_{i=1}^{n_{\rm comp}} a_k \boldsymbol{v}_k, \tag{20}$$

where $n_{\rm comp}$ is the number of leading components removed (typically four).

Figure 11 shows the three leading PCA components for a typical scan. For each component, its variation with time is shown for the duration of the scan, as well as its contribution to the overall variance for each feed. Although the contribution of even the leading PCA modes to the overall variance of a typical scan is on the level of single-digit percentages, recall that thermal noise will always dominate the variance for each scan and the spectral structure of even single-digit percentage PCA modes will surely dominate over the targeted CO signal, which is why this filter is important.

Figure 12 shows the frequency-channel-to-frequency-channel correlation matrix between all frequencies of all feeds before and after applying the PCA filter for a single scan. We see that most of the residual correlations between different feeds are removed in the PCA filter. A more extreme example, showing a case where the PCA mode dominates the variance of the data, is shown in Figure 13.

3.3.5. Masking

Sometimes individual frequency channels or groups of nearby frequency channels show artifacts, even after applying all the filters described above. This could manifest in a significant excess noise that is correlated in time, or in correlations between different frequency channels. We wish to mask these frequency channels so that their contribution does not contaminate the final results.

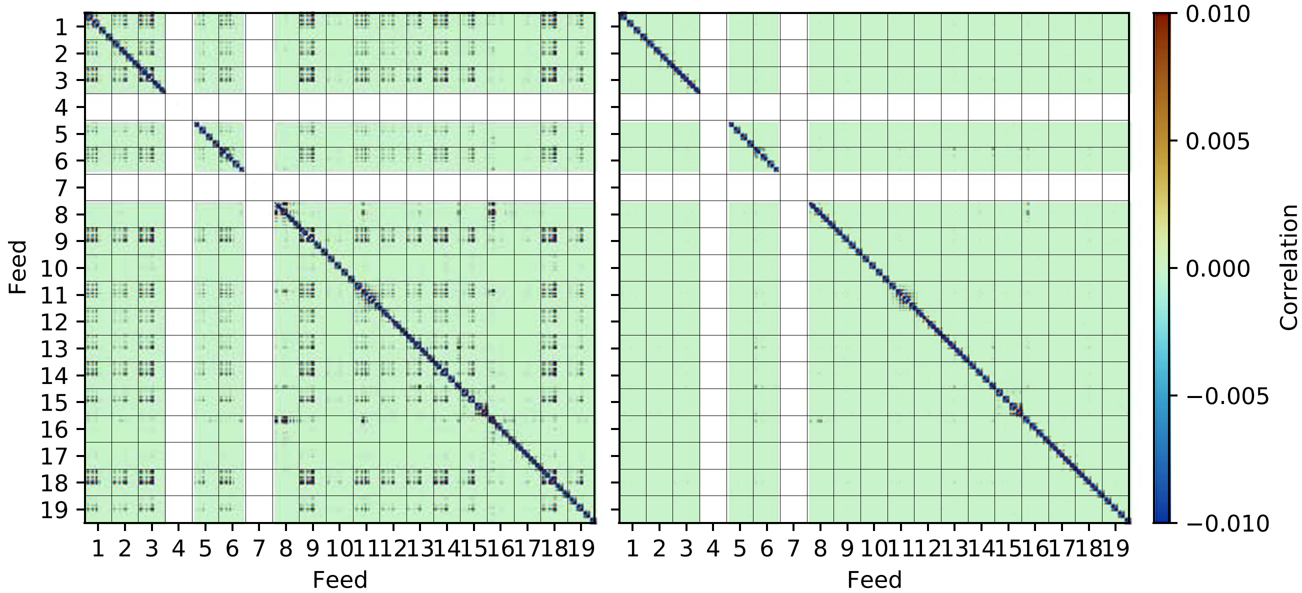


Figure 12. Comparison of channel–channel correlation matrices before (left panel) and after (right panel) applying the PCA filter.

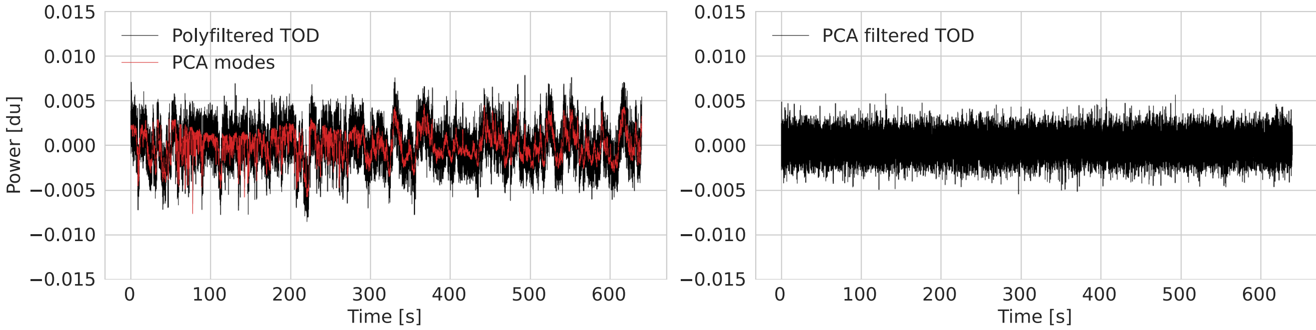


Figure 13. Effect of PCA filter on a "bad" scan, with unusually heavy weather or standing wave contributions.

To determine which frequencies should be masked, we first perform the polyfilters and PCA filters on a copy of the original data set. We then use two main approaches to identify individual or groups of frequency channels to be masked. The first approach uses the fact that the expected correlation between two independent Gaussian variables (for large $n_{\rm samp}$) is given by $1/\sqrt{n_{\rm samp}}$, where $n_{\rm samp}$ is the number of samples used to calculate the correlation. Thus, after accounting for the expected correlation induced by the polyfilter, we know the statistics describing *good* data and can identify bad data as deviations from these statistics. Specifically, we consider groups of elements within the frequency–frequency correlation matrix (either squares of different sizes or sets of columns) and compare the average absolute correlation within this group with the scatter expected from white noise alone. Any channel with an absolute correlation larger than 5σ is removed from further analysis.

Our second approach is to calculate a set of diagnostics for individual frequency channels, for instance, the average correlation of the channel in question to all the others in the same sideband, or the average absolute value of the same. We then compare the values of these diagnostics for the different channels and remove significant outliers.

In addition to these approaches, we also remove frequency channels heavily affected by aliasing. This typically corresponds to about 10% of the frequency channels, found at the edges of the bands. We mask all frequencies with a suppression of the aliased signal of less than 15 db. For more details on the aliasing effect, see Lamb et al. (2022). We also mask out individual frequency channels with very high system temperatures (above 80 K).

After the full mask has been determined, we apply the mask to the original (unfiltered) data set and repeat the filtering described above, but now only using the unmasked data. This prevents bad data from contaminating good data through the various nonlocal filters.

3.4. Calibration

With cleaned and co-added TOD in hand, the final step we need to perform at the TOD level is calibration, that is, assigning a noise temperature scale to the detector readout. From Equation (8), the overall noise level is proportional to $T_{\rm sys}$.

Ideally, in order to calibrate our instrument, we would put a load of a known temperature in front of the telescope and above the atmosphere and compare the measured output power with the output power measured with no load. A good approximation to this is to use an ambient temperature load that covers the receiver feed horn. Assuming that the telescope, the ground, and the atmosphere have the same physical temperature as the ambient load, the output power will be the same as if the load

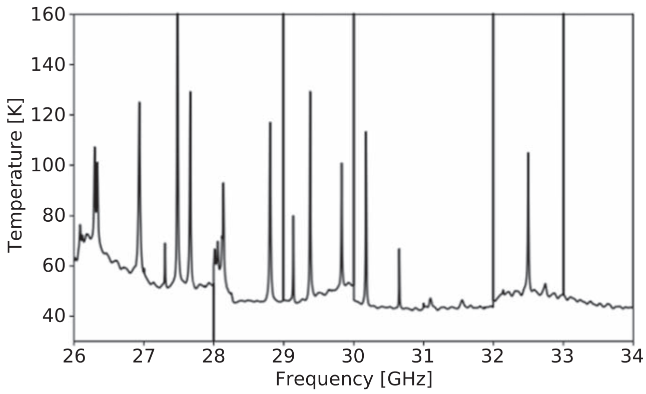


Figure 14. T_{sys} measurement from Feed 1 of obsID 15117 across the 4096 frequency channels.

was above the atmosphere (Penzias & Burrus 1973). Taking into account the vertical temperature profile and the distribution of the absorbing components in the atmosphere, the corrections are only a few percent for the relevant wavelengths.

In this approach we define the system temperature, T_{sys}, by

$$T_{sys} \equiv \frac{e^{\tau}}{\eta_{spill}}[T_{rx} + (1 - \eta_{spill})T_{gnd} + \eta_{spill}(1 - e^{-\tau})T_{atm} + \eta_{spill}e^{-\tau}T_{CMB}], \tag{21}$$

where T_{rx} is the noise temperature of the receiver, τ is the optical depth of the atmosphere, and $1 - \eta_{spill}$ is the fraction of the astrophysical signal lost to ground spillover. T_{gnd} and T_{atm} are the physical temperatures of the ground and the atmosphere, respectively, while T_{CMB} is the CMB monopole (we neglect other sky contributions). The overall factor of e^{τ}/η_{spill} converts from a system temperature defined at the receiver input to one defined outside the atmosphere. This definition ensures that

$$\Delta T_{sys} = \Delta T_{signal}, \tag{22}$$

meaning that a change ΔT_{signal} in the sky signal gives a corresponding change ΔT_{sys} in the system temperature. This definition makes the interpretation of our measurements easy and intuitive.

To measure the system temperature, we compare the readout when we have an ambient vane P_{amb} in front of the receiver and when we look at the cold sky P_{cold}. From Equations (7) and (21) we can estimate T_{sys} as

$$T_{sys} = \frac{T_{amb} - T_{CMB}}{P_{amb}/P_{cold} - 1}, \tag{23}$$

where T_{amb} is the ambient temperature and T_{CMB} is the cold sky temperature. We then multiply the data $d(\nu, t)$ by the T_{sys} measurement to go from (normalized) detector units to temperature

$$d(\nu, t)|_K = d(\nu, t)\langle T_{sys}(\nu, t)\rangle. \tag{24}$$

This method of calibration allows us to account for both the atmospheric absorption and ground spillover, without having to measure τ and η_{spill} themselves.

As both the atmosphere and the receiver gain vary over time, the measurements of T_{sys} vary over time. To get the most accurate estimation, we make use of the ambient vane at the beginning and end of each observation. We then interpolate the ambient load measurements to the time of each scan to calculate a value of T_{sys}, which is done for each feed and frequency of each scan in `l2gen`. Figure 14 shows a typical example of how the estimated T_{sys} looks for a single obsID, as a function of frequency. The temperature usually behaves as a relatively smooth function, with large spikes at specific frequencies (for more details see Lamb et al. 2022). To see the distribution of sideband-averaged system temperature for all Season 1 observations, we refer the reader to Figure 21 in the Appendix.

A challenge with this calibration method is that we are calibrating the total power of the instrument, integrated out to about 90°, rather than just the power in the main beam. As we are interested in structures at small angular scales, some of the total power is essentially lost, with the details depending on the structure of the beam and the scales of interest. In the power spectrum analysis (Ihle et al. 2022) we take this into account by using a beam transfer function, calibrated on measurements of astrophysical calibration sources (Rennie et al. 2022).

3.5. Downsampling

Until now, all steps have been performed at full frequency resolution, i.e., 1024 channels per sideband or 2 MHz channel bandwidth. For mapmaking purposes, however, we typically do not require such high resolution, as the intrinsic line width of the CO signal limits the amount of information at small line-of-sight scales (Chung et al. 2021). To save both memory and computing time, we therefore co-add several neighboring high-resolution frequency channels (usually 16, corresponding to a final bandwidth of 31.25 MHz) into a single low-resolution channel using inverse variance noise weighting.

3.6. Mapmaking

After the main data selection step (described in Section 4), the last step in the pipeline is mapmaking, which is implemented in a code called `tod2comap`. This reads in cleaned TOD and pointing information, applies a high-pass filter, and produces temperature sky maps for each frequency channel. The high-pass filter removes structures on long timescales in the TOD. This is done by Fourier-transforming the TOD and removing the part with frequency below a set value, typically 0.02 Hz, before transforming back to TOD.

Ideally, the TOD can be written as a sum of the signal $\boldsymbol{s}$ and the noise $\boldsymbol{n}$,

$$\boldsymbol{d} = \boldsymbol{P}\boldsymbol{s} + \boldsymbol{n}, \tag{25}$$

where $\boldsymbol{P}$ is the pointing matrix, which connects each time sample to a pixel on the sky. Our goal is to estimate $\boldsymbol{s}$ given $\boldsymbol{d}$. Assuming that the noise is Gaussian distributed with a time-domain covariance matrix $\boldsymbol{N}$, the log-likelihood function corresponding to Equation (25) may be written as

$$\log\mathcal{L} \propto (\boldsymbol{d} - \boldsymbol{P}\boldsymbol{s})^T\boldsymbol{N}^{-1}(\boldsymbol{d} - \boldsymbol{P}\boldsymbol{s}). \tag{26}$$

Setting the derivative of this log-likelihood to 0, we obtain the standard mapmaker equation,

$$\hat{\boldsymbol{s}} = (\boldsymbol{P}^T\boldsymbol{N}^{-1}\boldsymbol{P})^{-1}\boldsymbol{P}^T\boldsymbol{N}^{-1}\boldsymbol{d}. \tag{27}$$

As discussed above, the COMAP noise after filtering is very close to white, and this implies that $\mathbf{N}$ may be approximated as diagonal.[18] In that case, Equation (27) may be solved explicitly and independently for each pixel p as follows:

$$\hat{s}_p = \frac{\sum_{t\in p}\sigma_t^{-2} d_t}{\sum_{t\in p}\sigma_t^{-2}}. \tag{28}$$

Here σ_t is the noise standard deviation of sample t, and samples with lower noise are thus weighted more strongly than the samples with higher noise. The corresponding map-domain noise standard deviation is given by

$$\sigma_p = \left(\sum_{t\in p} \frac{1}{\sigma_t^2}\right)^{-1/2}. \tag{29}$$

We perform this mapmaking procedure separately for each frequency channel.

4. Data Selection, Observation Efficiency

As we will show in Section 6, most of our filtered time streams are dominated by white noise. However, this does not necessarily imply that they are free from systematic errors to a level required for scientific analysis. On the contrary, many effects may only be discovered when co-adding both over time and frequency.

The main challenge for data selection is to identify and remove data contaminated by systematic errors. It is preferable to remove bad data at the earliest stage possible, before they are co-added with clean data. However, co-adding data also reduces the noise, making it easier to identify systematic effects at a later stage. For this reason, since we cannot detect all systematic errors during the low-level filtering and masking, we go through several stages of data selection, throughout the data analysis pipeline.

In addition to the frequency masking described in Section 3.3.5, we also apply cuts based on statistics calculated for each sideband of each feed and scan. These statistics allow us to find patterns and correlations at levels far below the noise level of an individual scan.

4.1. Data Losses and Efficiency

In order to quantify the overall data efficiency, $E_{\rm data}$, i.e., the fraction of raw data we use for the final power spectrum estimates, we summarize the different stages at which data are rejected:

1. No data from feeds 4, 6, and 7 were used for the final analysis. Two of these feeds (4 and 7) were used for engineering tests and did not produce useful data, while large systematic errors were visible in the low-level data for one of them (feed 6). We denote the fraction of data lost by rejecting data from these feeds by $L_{\rm feed}$.
2. As described in Section 3.3.5, during the low-level data filtering (in `l2gen`), we mask bad, outlier or aliased frequency channels. This means that we lose some fraction of the data, denoted by $L_{\rm freq}$.
3. During the first year of observations, we took a large amount of data at elevations above 65° and below 35°. Since we now know that these data are very susceptible to ground contamination (via the main beam and sidelobes), we do not use these data in our results. The fraction of data lost in this elevation cut is denoted by $L_{\rm el}$.
4. The main data selection stage (in `accept_mod`) consists of two main parts. First, we go through all the level 2 files and gather or calculate a long list of diagnostic statistics for each scan, for example, the time of day, the average system temperature, and a range of goodness-of-fit statistics. Other examples include whether or not the Sun or the Moon is in one of our main sidelobes, the measured noise properties, machine-learning-derived weather predictions (Rasmussen 2020), and various parameters from the low-level filters. In total there are 77 different such statistics in the database. Each of these statistics is calculated for each sideband of each feed for each scan. The next part of the process consists of defining, for most of these statistics, an allowed range of values. Using the full database, together with the allowed range of values, we make an *accept list*, which determines which data are accepted and which are rejected at this stage. We denote the fraction of the data that are lost at this stage (excepting the `ps_chi2` statistics mentioned next) by $L_{\rm stats}$.
5. Of the 77 statistics described above, one group is of particular importance. We take the filtered data from a single scan and make a single 3D map using the data from a single sideband of a single feed. We can then divide this map by the white-noise variance map and calculate the spherically averaged pseudo-auto-spectrum for this single sideband. This pseudo-auto-spectrum can then be compared to what we would expect from white noise and the different data points combined into a single χ^2 goodness-of-fit statistic that we call the `ps_chi2`. We make different versions of the `ps_chi2` statistic, for example, by combining the four sidebands of each feed into a single map, by combining all feeds into a single map for a single scan, or by combining maps from all scans in a single observation (roughly 60 minutes of data) into a single map, calculating a separate `ps_chi2` statistic for each of the different maps. We then set thresholds for acceptable deviations from the white-noise expectations and reject data that show large excesses in these power spectrum statistics. We denote the fraction of data that are lost at this stage by $L_{\chi^2_{P(k)}}$.
6. During the final stages of the main power spectrum estimation, we calculate the cross-spectra for data from different feeds and different data splits (for more details on this see the companion paper Ihle et al. 2022). For each of these spectra we calculate χ^2 statistics that are used to accept or reject the spectra in the final results. This allowed us to identify problems associated with specific feeds. For example, feed 8 had a known problem with the LNA, and almost all cross-spectra involving this feed had a high χ^2 statistic; we therefore removed all spectra involving this feed at this step. For Field 1 we also found clear excesses in several spectra involving feeds 16 and 17 from the low-elevation data set; all these

[18] This is not strictly correct for long timescales. As such, the current mapmaker is statistically slightly suboptimal, and the resulting transfer function is lower than strictly necessary. Future implementations of the COMAP mapmaker will therefore instead rely on well-established destriping or maximum likelihood algorithms, which are often able to recover slightly more large-scale information than a binning mapmaker.

Table 3
Losses and Overall Data Efficiencies for the First-season Data (Columns (1)–(4)) and an Optimistic Projection for Years 2–5 (Column (5))

	Field 1	Field 2	Field 3	All Fields	Projection
$L_{\rm feed}$	15.8%	15.8%	15.8%	15.8%	0.0%
$L_{\rm freq}$	26.7%	28.1%	26.7%	27.2%	15.0%
$L_{\rm el}$	7.3%	31.2%	29.5%	24.4%	0.0%
$L_{\rm stats}$	47.4%	35.7%	44.9%	42.6%	35.0%
$L_{\chi^2_{P(k)}}$	20.9%	22.3%	40.0%	27.8%	20.0%
$L_{\chi^2_{C(k)}}$ [a]	24.6%	78.8%	39.6%	47.6%	10.0%
$E_{\rm data}$	18.0%	4.42%	8.70%	9.50%	39.8%

Notes. Losses, L, denote the fraction of data lost at each step. Here $E_{\rm data}$ is the product of the factors $(1 - L)$ for all the losses in the rows above.
[a] These are the losses for the CES cross-spectra, which are the only ones we ended up using in the final results. The corresponding losses for the Lissajous data are given in Ihle et al. (2022).

spectra were also removed from Field 1. We also cut any spectrum with a χ^2 above 5σ. We denote the fraction of data that is lost at this stage by $L_{\chi^2_{C(k)}}$.

7. Power spectra formed from the data taken using the Lissajous scanning strategy (which we used for about half of the observations in the first season) showed strong large-scale excess power, potentially due to ground contamination, which is more easily removed from CES scans. (This excess power is most clearly quantified in Table 1 of Ihle et al. 2022.) For this reason, we did not include any of the Lissajous data in our final science results. We denote the fraction of the final data using CES scans as $E_{\rm scan}$.
8. Finally, there are periods in time where the telescope, for whatever reason, is not observing the main science CO fields. We denote the fraction of time that we are taking CO data by $E_{\rm obs}$. For each of our three individual CO fields we define the observation efficiency as the total time the field was observed multiplied by 3 and divided by the total period over which the observations were taken.

4.2. Future Sensitivity Projections

Table 3 shows the data lost at different stages of data selection, as well as an optimistic projection for how these values could change in the future. As we can see, a large fraction of the data are lost in the final stage of cuts (based on the $\chi^2_{C(k)}$ statistics), indicating that there are systematic errors that are not being identified in earlier steps. By understanding the origin of these errors and removing them at an earlier stage, there is the potential to significantly increase the amount of data available for analysis.

As mentioned, in the case of data taken using Lissajous scans (which corresponds to about half of the total obtained in Season 1), there is a clear excess in the final power spectrum; for this reason, these data were not used for our science results. For data taken using CES scans, the $\chi^2_{C(k)}$ cut produces spectra for Fields 1 and 2 that are consistent with white noise. For Field 3 we needed to apply a more restrictive set of limits on the various statistics and $\chi^2_{P(k)}$, which we believe to be related to an increased level of ground contamination compared to the other two fields.

With experience of Season 1 in hand, we are working on building the second-generation COMAP pipeline, including improved ground modeling, mapmaking, and real-time continuum filtering and calibration, based on the lessons learned from the first-generation data analysis and our improved understanding of the data. We have also altered our observing strategy and corrected hardware problems. For some of these improvements, quantifying the resulting increase in sensitivity is somewhat speculative. However, the combined effect of all of these changes is likely to be significant, and we will therefore attempt to systematically estimate the amount of usable data that will be available after 5 yr, taking these changes into account, and how this transfers to our power spectrum limits. This will allow us to compare our estimated 5 yr sensitivity to signal models (Chung et al. 2022) and serve as a benchmark against which we can compare our future progress.

Below we discuss the improvements we think we potentially can achieve in each factor from Table 3.

1. At the end of Season 1, the receiver was removed from the telescope for maintenance. Feeds 4 and 7 were switched from engineering testing to science operations mode, while problems with feed 6 were repaired. In subsequent observing seasons, we therefore hope to keep $L_{\rm feed}$ close to 0%.
2. We plan to increase the clock frequency of the analog-to-digital converters in the ROACH-2 back end. This will reduce the number of frequency channels removed owing to aliasing and, coupled with improvements in the filtering and a more stable system, should improve $L_{\rm freq}$ significantly.
3. Regarding $L_{\rm el}$, we are no longer observing above 65° and below 35° in elevation, so we should not lose any more data to this elevation cut in subsequent seasons.
4. With a careful study of the effect of relaxing the current conservative limits on the various statistics, we believe that there is some scope to reduce $L_{\rm stats}$ by identifying which cuts are important and relaxing the others.
5. As we have not yet spent time to fine-tune the allowed limits on these statistics, we believe that it will be possible to significantly improve $L_{\chi^2_{P(k)}}$ if we choose the limits more carefully.
6. By identifying data affected by systematic errors at an earlier stage in the pipeline and by splitting the data into more pieces for the cross-correlation, we expect to be able to significantly reduce $L_{\chi^2_{C(k)}}$ (especially for Fields 2 and 3). Field 1 shows that it is possible to reduce the losses at this step significantly. For Field 1, after we removed all spectra involving feed 8 and all spectra involving feeds 16 and 17 from the low-elevation data set by hand, the automatic $\chi^2_{C(k)}$ cut at 5σ accepts all but one of the remaining 182 cross-spectra, indicating that the remaining data are very clean.
7. After Season 1 we no longer use Lissajous-type scans, which means that $E_{\rm scan} = 100\%$ for all subsequent seasons.
8. During Season 1, we addressed the main instrumental and operational issues that decreased $E_{\rm obs}$, as well as instituting weekly maintenance checks; we expect to achieve close to the maximum efficiency of 82.5% (based on the total time our CO fields are within 35°–65°

elevation) for future seasons. We will therefore assume an observation efficiency of 75% for the future, a large improvement over the values of 36.8%, 52.9%, and 53.2% for Fields 1, 2 and 3, respectively, obtained during the first season.

Based on these considerations, we make an estimate of our future data efficiency, as shown in Table 3 (the rightmost column). We further assume that we can estimate the future (5 yr) power spectrum sensitivity by simply scaling the current sensitivity by an overall factor, D,

$$\sigma^{5\,\mathrm{yr}} = \frac{\sigma^{\mathrm{S1}}}{D^{5\,\mathrm{yr}}}. \tag{30}$$

We define the total efficiency for Season 1 as $E_{\mathrm{tot}}^{\mathrm{S1}} \equiv E_{\mathrm{scan}}^{S1} E_{\mathrm{data}}^{S1} E_{\mathrm{obs}}^{S1}$, where $^{\mathrm{S1}}$ denotes the quantity from the first season. The fraction of data using the CES scan during the first season, E_{scan}^{S1}, was given by 51.7%, 55.6%, and 34.3% for Fields 1, 2, and 3, respectively. We assume for the forecast that by splitting the data into more parts (than the two we are currently using), we can improve the sensitivity of the cross-spectrum estimator (see Ihle et al. 2022 for more details) by a factor of $E_{\mathrm{split}} = 1.3$ (the asymptotic limit as $N_{\mathrm{split}} \to \infty$ is $\sqrt{2}$).

Using the total duration of the season 1 observing campaign, $T^{S1} = 440$ days, we base our forecast on the performance of our best field (Field 1) and find the factor, $D^{5\ \mathrm{yr}}$, needed to estimate the 5 yr sensitivity, assuming that we can make all three of our fields perform as well as Field 1. We also assume that we can improve the transfer function by at least 10% on average by improved filtering and mapmaking, giving us an extra overall factor $E_{\mathrm{TF}} = 1.1$. We then get

$$D^{5\,\mathrm{yr}} \equiv \frac{T^{S1}E_{\mathrm{tot}}^{\mathrm{S1}}/\sqrt{3} + (5\cdot 365 - T^{S1})E_{\mathrm{tot}}}{T^{S1}E_{\mathrm{tot}}^{\mathrm{S1}}/\sqrt{3}} E_{\mathrm{split}} E_{\mathrm{TF}}, \tag{31}$$

where $E_{\mathrm{tot}} \equiv E_{\mathrm{data}}^{\mathrm{proj}} E_{\mathrm{obs}}^{\mathrm{proj}} = 29.8\%$, and where the $\sqrt{3}$ comes from the fact that we are extrapolating the current sensitivity of Field 1 to all three fields. Here the superscript "proj" denotes the previously discussed future projections. Inserting the values, we find $D^{5\ \mathrm{yr}} = 69$. Discussion of the current upper limit and the 5 yr power spectrum sensitivity forecast can be found in the companion paper Chung et al. (2022).

The preceding number should be interpreted as a reasonably optimistic order-of-magnitude estimate and does not take into account new systematic errors that may be revealed with any increase in sensitivity. Such effects will surely require a revision of our existing filtering and data selection procedures.

5. Signal Loss and the Pipeline Transfer Function

5.1. The Pipeline Transfer Function

The main summary statistic we use to estimate the CO signal is the power spectrum

$$P(\boldsymbol{k}) = \frac{\langle |f_{\boldsymbol{k}}|^2 \rangle}{n_x n_y n_z} V_{\mathrm{vox}}. \tag{32}$$

This is extracted from the temperature sky maps by, first, computing the 3D Fourier transform of the maps; binning the squared Fourier coefficients according to the wavenumber, $\boldsymbol{k}$; and averaging over all contributions to a given $\boldsymbol{k}$-bin. Finally, they are multiplied by the comoving voxel volume, V_{vox}, and divided by the total number of voxels, $n_x n_y n_z$. Note that each voxel is inverse variance weighted by σ_p^{-2} as given by Equation (29) before computing the Fourier transform, meaning that we are calculating the pseudo-spectrum, or the spectrum of the inverse variance noise weighted map, rather than a regular auto-spectrum. It is therefore important to keep in mind that the pipeline transfer function deduced from pseudo-spectra, which is what will be discussed in this section, will be similarly biased as the pseudo-spectra themselves. We make no effort here to account for or undo the "mode mixing" resulting from the noise weighting, but we leave the discussion about this effect and power spectrum methods in general to Ihle et al. (2022).

As the raw data pass through our filtering and mapmaking procedures, some of the signal is typically lost at each stage, and the maps described in Section 3.6 are therefore biased. In order to estimate and correct for this bias at each scale, $\boldsymbol{k}$, we need to estimate the so-called pipeline transfer function, which is simply defined as the power spectrum ratio between the recovered and original signal.

We can estimate this transfer function by adding a signal-only simulation to a pure noise TOD and then comparing the combined signal-plus-noise simulation output to the true signal-only input. We adopt the raw COMAP TOD as a model for the noise, which in power units are denoted by P_{N}. The signal-only contribution is produced by scanning a precomputed 3D simulation of brightness temperature (using the fiducial model in Chung et al. 2022), $T_{\mathrm{sim}}(p)$, with the telescope pointing, and we label this P_{S}. We then add these together in power units,

$$P_{\mathrm{S+N}} = P_{\mathrm{N}} + P_{\mathrm{S}} = k_{\mathrm{B}} G \Delta\nu T_{\mathrm{sys}}\left(1 + \frac{T_{\mathrm{sim}}}{T_{\mathrm{sys}}}\right). \tag{33}$$

If this is done at each time step (and pointing) of the raw data, then we can construct a simulated TOD $P_{\mathrm{S+N}}$ simply by adding the temperature of the simulated cube of signal at any given frequency channel along the line of sight. This signal-plus-noise TOD can then be sent through our whole low-level data analysis pipeline, like a regular TOD. We only need to make sure that the same frequency masking is applied to the TODs with or without added signal, to make it a fair comparison. We then separately generate 3D voxel maps from P_{S}, P_{N}, and $P_{\mathrm{S+N}}$, and from these we compute corresponding 3D pseudo-auto-spectra $P_{\mathrm{S}}(\boldsymbol{k})$, $P_{\mathrm{N}}(\boldsymbol{k})$, and $P_{\mathrm{S+N}}(\boldsymbol{k})$, following the above procedure. Based on these three spectra, we can finally estimate a scale-dependent transfer function $T(\boldsymbol{k})$ as

$$T(\boldsymbol{k}) = \frac{P_{\mathrm{S+N}}(\boldsymbol{k}) - P_{\mathrm{N}}(\boldsymbol{k})}{P_{\mathrm{S}}(\boldsymbol{k})}. \tag{34}$$

Noting that the pipeline filters have very different impact in the angular and frequency directions, it is useful to decompose $\boldsymbol{k}$ into parallel (line-of-sight) modes, $k_{\parallel} \equiv |k_z|$, and the perpendicular (angular) modes, $k_{\perp} \equiv \sqrt{k_x^2 + k_y^2}$. This is the version of the transfer function we use for the main science analysis. However, for simplicity of visualization we will here show several results for the 1D (spherically averaged) version of the transfer function, in addition to the 2D (cylindrically averaged) one.

Another thing to note is that the signal used to estimate the transfer functions is boosted, compared to the theoretical

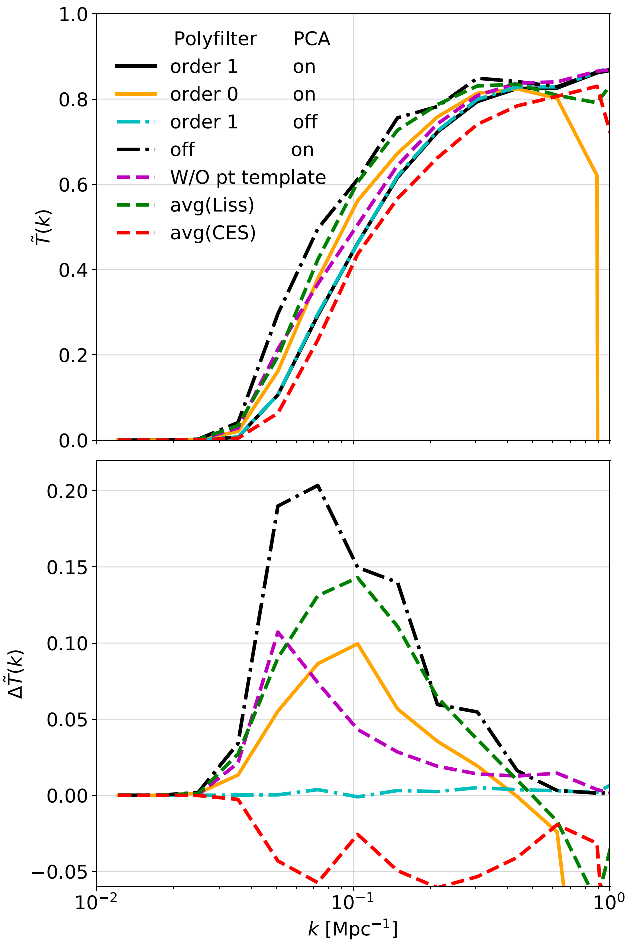


Figure 15. Top panel: 1D transfer functions, $T(k)$, for different filter options and scanning modes as a function of scale, k. The default combination used in the COMAP pipeline is shown as a solid black line. Bottom panel: difference between the various filter and scanning options and the default configuration.

model, so that its peak temperature is around 2 K, in order to make it easily detectable given the level of noise TOD used. As a result, the signal-to-noise ratio in the simulated data will be several orders of magnitude higher than in the raw data, as the actual CO signal from a raw observation will be completely dominated by noise when only using a few hours of observation.

5.2. The Effect of Individual Filters on the Transfer Function

First, to understand the impact of the various filters in terms of signal loss, we estimate 1D transfer functions for a range of different pipeline configurations. Specifically, we analyze six obsIDs (three CES and three Lissajous obsIDs, both observations of Field 3), where we consider different combinations of PCA and polyfilter, enabling or disabling each filter in turn. For the polyfilter, we additionally consider two cases, namely, a constant fitting term or a linear fitting function. Here, we only wish to illustrate the effects of each filter on the measured signal, and so we have simply added signal to the raw data using a single simulation realization. This ensures that the effects seen in each filter combination are not due to any differences between realizations. However, it will be necessary in future analyses, when even the smallest systematic effects become important, to average the transfer function estimates resulting from several different realizations. The results from these calculations are summarized in Figure 15. The black solid line shows the default pipeline configuration.

One can see that the default settings, i.e., a first-order polyfilter and PCA filtering turned on, yield almost the same transfer function as the case where the PCA filter is turned off. The PCA filter is not expected to remove much of the actual input signal, as it only removes the components of the TOD that are the most correlated over all frequencies and feeds, thus potentially removing only the structures of the input signal that are common over the entire survey volume observed at any given time.

When it comes to the case with a zeroth-order polyfilter or with the polyfilter turned completely off, there are, however, large differences seen from the results using the default settings. Using a zeroth- or first-order polyfilter, a considerable fraction of the input signal is removed by the pipeline on scales above $k \sim 0.04$ Mpc^{-1}. We see that the zeroth-ordered polyfilter yields a similar result to that without the polyfilter near the peak regions of the transfer functions; however, a nonnegligible portion of signal from $k \sim 0.04$ Mpc^{-1} up to the peak region is taken out when turning on the zeroth-order polyfiltering. The low transfer efficiency on low k for any of the shown filter combinations is due to the limited area covered in each scan, the high-pass filter imposed in the mapmaker, and the polynomial filter in frequency.

If we turn off the subtraction of the Az/El templates, we can also see in Figure 15 that more signal is let through the pipeline on scales $k \leqslant 0.3$ Mpc^{-1}. The effect of the Az/El template subtraction is, however, especially noticeable on scales $k < 0.1$ Mpc^{-1}, which is expected, as the structures in the power spectra removed by the Az/El templates are of a larger scale in the pixel domain.

Note also that when computing these transfer functions for different filter combinations, we used the combined maps of three obsIDs of Lissajous-type scans and three with CES. However, we found that there were significant differences between the transfer functions from a Lissajous and CES scan type and have therefore also included the average of the three transfer functions of each type in Figure 15. As one can see, the Lissajous scan type results in a transfer function that is larger on most scales, which probably is a result of the Lissajous scan covering a larger area in a single scan compared to the CES. The Lissajous scans, as opposed to the ones with CES, also seem to result in a transfer function that drops a bit down from its peak at high k. The reason for this difference is not yet fully understood at this point.

In general, the estimates of the transfer function break down at high $k_\perp$ owing to $P_S(\boldsymbol{k})$ going to zero in Equation (34), so we see some large random fluctuations here, but this is not a problem because the instrumental beam means that we have basically no sensitivity to these modes anyway.

When looking at the 2D version of the transfer function, as shown in Figure 16, the effects of the polyfilter on the transfer function become more evident because we can then distinguish between what is happening in the angular directions ($k_\perp$) and the spectral dimension ($k_\parallel$). As the polyfilter is designed to remove the $1/f$-noise and continuum foreground emission along the frequency dimension on each sideband, we expect the changes in the transfer function to be most visible in the large line-of-sight scales $k_\parallel$. This is indeed what is seen in the difference $\Delta T(k)$ between the transfer functions without and

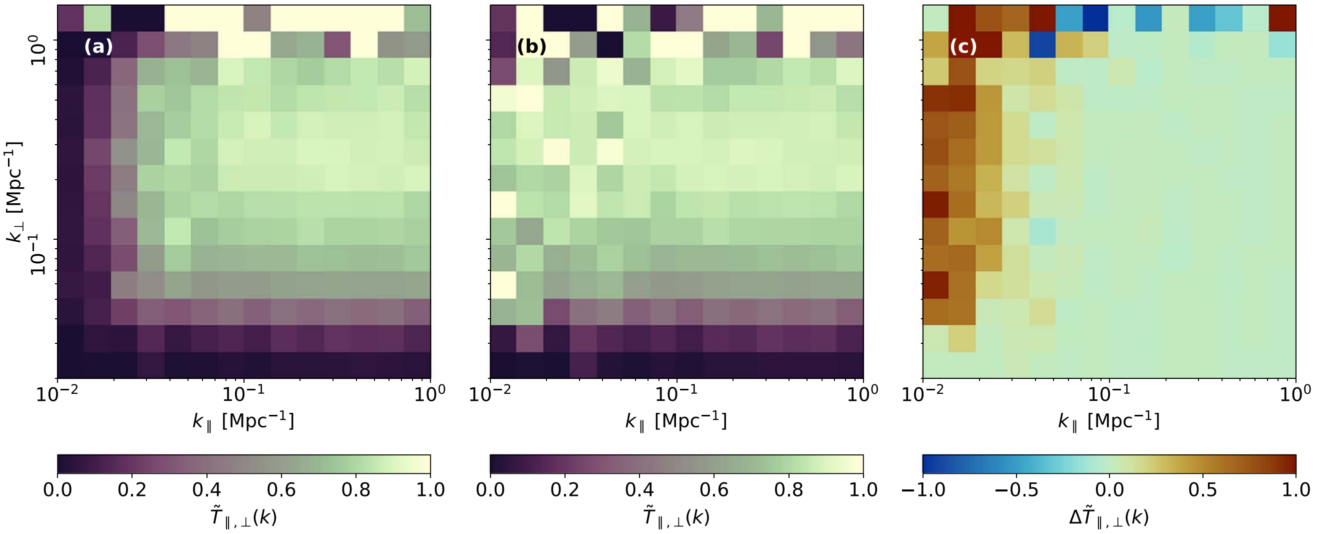


Figure 16. Comparison of 2D transfer function estimates with (left panel) and without (middle panel) the polyfilter. Right panel: difference between the two previous cases.

with a first-order polyfilter in Figure 16 for low $k_\parallel$, where we note a 50%–90% relative loss in power when using a first-order polyfilter. Meanwhile, on all other scales the difference ΔT is left mostly unchanged.

Lastly, we emphasize that the transfer functions presented here are meant only to graphically illustrate the effects of our filtering on the signal. The transfer functions applied to the final pseudo-power spectrum, to compensate for the bias introduced by the filtering, are cylindrically averaged and are based on roughly 63 hr of observations. For more details on this see Ihle et al. (2022).

6. Noise Characterization and Removal of Correlated Noise

In this section we describe our noise characterization model and demonstrate how effectively the pipeline removes correlated noise from the data. One of the most important functions of the COMAP data pipeline is the removal of correlated noise. By correlated noise, we mean noise that is correlated in time. The fact that all known sources of correlated noise in our system also produce noise that is correlated across feeds or frequencies, gives us powerful leverage to remove correlated noise from our data. As the CO brightness temperature is many orders of magnitudes below the telescope system temperature, any significant deviations from a white-noise spectrum in our filtered data must be due to residual correlated noise or another systematic effect. This means that in order to quantify the presence of correlated noise, we can look at the deviation from white noise.

We can often approximate the correlated noise using a spectral density of the form

$$N(f) = \sigma_0^2\left(1 + \left(\frac{f}{f_{\rm knee}}\right)^\alpha\right), \tag{35}$$

where σ_0 is the white-noise level.

The second term in Equation (35) is known as $1/f$ noise,[19] which is characterized by a knee frequency $f_{\rm knee}$, representing the transition frequency between the flat white noise and the sloped $1/f$ noise, and the exponent α, giving the slope of the spectral density in the $1/f$-dominated regime. The white-noise level is estimated by calculating the variance between neighboring samples in the TOD, as

$$\sigma_0 = \sqrt{\frac{\mathrm{Var}(d_i - d_{i-1})}{2}}. \tag{36}$$

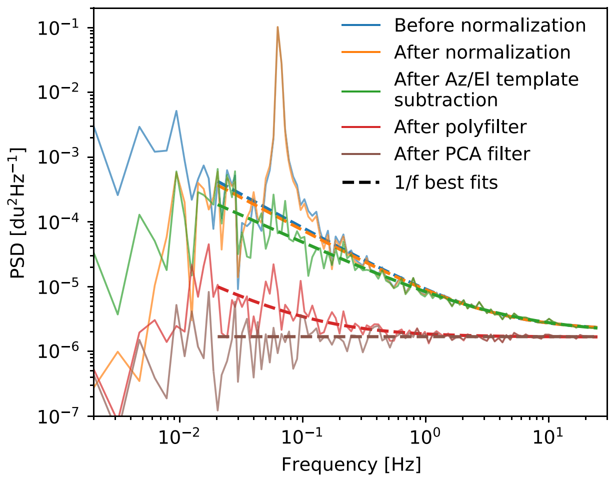


Figure 17. Power spectral distribution of a single scan from a 31.25 MHz band around 28.2 GHz at different stages in the pipeline, with $1/f$ noise curves fitted. The power spectral distribution is binned with logarithmic bin sizes toward higher frequencies for clarity. Lower frequencies have been excluded from the fit, as these scales are greatly suppressed at the normalization stage.

Figure 17 shows the $1/f$ behavior of the TOD throughout different steps in the pipeline and clearly demonstrates the effect of each filter. The normalization step heavily suppresses the low-frequency end of the spectrum. The Az/El template knocks out the strong $\sim$0.7 Hz correlation caused by the Lissajous scanning strategy. The polyfilter significantly reduces the noise power across the entire power spectrum, even lowering the white-noise limit. This is possible because even though the white noise is uncorrelated in time, parts of it are still correlated in frequency. Finally, the PCA filter further

[19] Keep in mind that f refers to the temporal frequency of the time-ordered signal, not the observed photon frequencies, which we consistently refer to as ν.

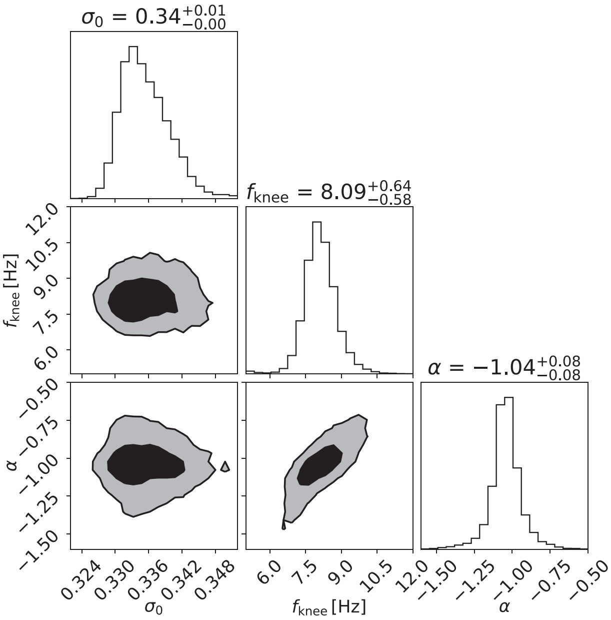


Figure 18. Distribution of noise parameters σ_0, $f_{\rm knee}$, and α for the c_0 coefficient of the polyfilter. All available scans of feed 1, sideband A:LSB were used.

reduces the noise left over by the polyfilter. By the end of the pipeline, the TOD is almost completely dominated by white noise. It should be noted that while the polyfilter typically suppresses much more noise power than the PCA filter in an average scan, this is not always the case. In scans with significant contamination (like standing waves or bad weather), the PCA filter may suppress even more noise power than the polyfilter. An extreme example of this is shown in Figure 13.

6.1. Polyfilter Noise Properties

As discussed in Section 3.3.3, the polyfilter involves fitting and subtracting a low-order polynomial in frequency space from each sideband at each individual time step. The polyfilter is the first filter targeting correlated noise except on the very largest timescales, and the resulting coefficients are therefore highly informative regarding the noise properties of the data. In the current analysis setup, we only use a first-order polynomial filter, such that each time step of each sideband is associated with two coefficients, c_0 and c_1. These coefficients, treated as functions of time, turn out to have $1/f$-like power spectra. Figure 18 shows the distribution of noise parameters of $1/f$ fits performed on c_0 for all available scans of the A:LSB sideband of feed 1. As discussed in Section 2.4, the correlated noise common to each sideband is mostly dominated by gain fluctuations of the individual LNAs at each feed. We therefore expect, and find, that each feed has its own characteristic noise properties. Since we can use the polyfilter to remove this correlated noise, the individual noise properties of the different feeds are less important when measuring the CO line emission than if we were measuring continuum sources, in which case these properties would become crucial.

6.2. Goodness of Fit, χ^2 Test

The main goal of our pipeline is to remove both correlated noise and continuum foregrounds, while leaving as much as possible of the CO line intensity signal intact. In the ideal case, and assuming that the cosmological CO signal is so weak that it cannot be measured in a single scan, our cleaned TOD should therefore be described by white noise alone. We therefore need statistics to measure potential deviations from white noise. We use a standard χ^2 statistic per scan for this purpose, defined as follows:

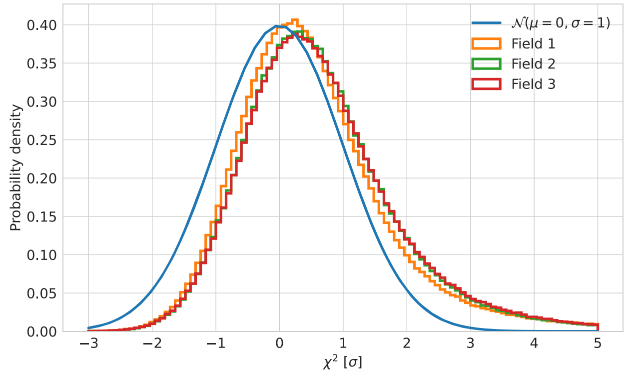


Figure 19. χ^2 distributions of filtered data for the three main fields, with a standard normal distribution for comparison.

$$\chi^2 = \frac{\sum_{i=0}^{N}\left(\frac{d_i}{\sigma_0}\right)^2 - N}{\sqrt{2N}}. \tag{37}$$

Here d_i are the N samples of the scan, and σ_0 is the white-noise level defined in Equation (36). For a perfect white-noise TOD, we expect $\chi^2 \sim \mathcal{N}(\mu = 0, \sigma = 1)$.

Figure 19 shows the χ^2 distribution for all the scans in the first observation season, comprising about 5000 hr of observations, divided by observational field. Here we have combined all the data points for each sideband, such that $N = n_{\rm samp} \cdot n_{\rm freq}$, where $n_{\rm samp}$ is the number of samples in time (typically $n_{\rm samp}$ = 10–20,000) and $n_{\rm freq} = 64$ is the number of frequencies per sideband. As seen in Figure 19, the data are indeed very close to white noise, with only a small shift and a positive tail. We also note that the Field 1 field outperforms the two other fields by a small margin. Given that the number of samples, $N \sim 10^5$, going into each of the χ^2 values in this histogram is so large, a mean bias of less than 1σ per scan corresponds to an excess variance (from correlated noise or other systematic effects) of less than about 0.5%. This is remarkable, since we have not imposed a high-pass filter to remove the correlated noise; nevertheless, our filters (mostly the polyfilter and the PCA filter) are able to remove it almost perfectly by using the fact that it is correlated between different frequencies. Since correlated noise integrates down as we add together independent observations, this means that the correlated noise will always be dominated by the white noise. Other systematic effects, however, may not be entirely independent between different observations and will thus not necessarily integrate down as fast when we combine different observations. Such systematic effects can thus lead to problems that are not visible at this stage of the analysis.

6.3. Maps

Figure 20 shows a single-frequency map, from each of our three fields, based on the data from the first season of observations. This data set results in a sensitivity of a few tens

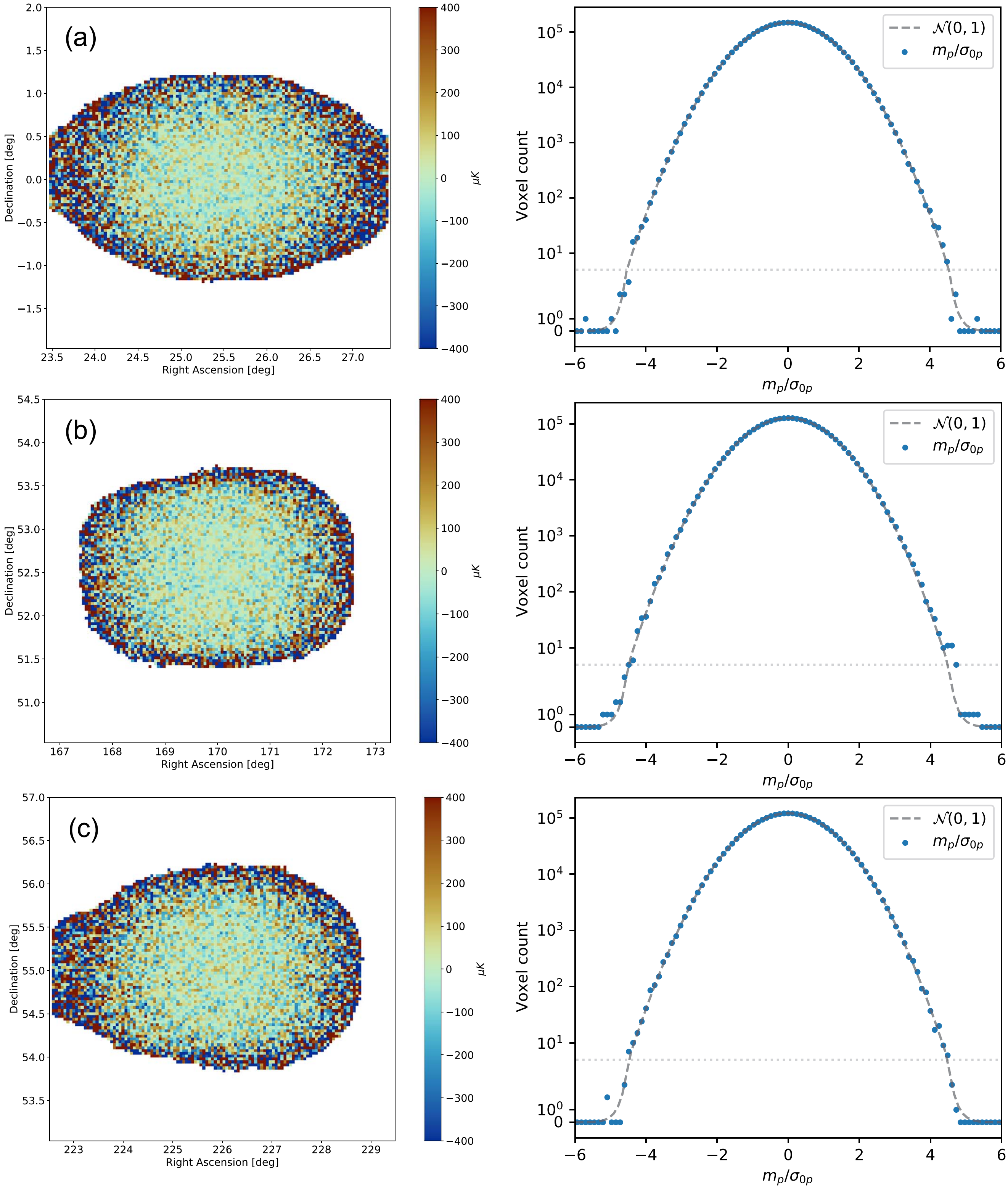


Figure 20. Co-added COMAP single 31.25 MHz frequency channel maps with a central frequency of 28.9 GHz (left) and voxel histograms of the map voxels divided by their corresponding white-noise level for all 3D voxels (right) for (a) Field 1, (b) Field 2, and (c) Field 3. Regions that either are not observed by the telescope or have a noise level $\sigma_0 > 1000\ \mu$K are masked out in the plotted maps. Note that in the voxel histogram we use a linear y-axis below a voxel count of 5 and a logarithmic one above.

of μK per 2×2 arcmin^2 pixel for a single 31.25 MHz channel. At least at a visual level, the maps appear largely dominated by white noise. We see that each field only has significant coverage within roughly a 2×2 deg^2 area on the sky. The right panel of the figure shows histograms of all the map voxels, m_p (Equation (28)), divided by their corresponding white-noise level, σ_{0p} (Equation (29)). Overplotted is what is expected from a unit normal variable. The histogram shows that we are extremely close to white noise even out to the far tails of the distribution. While this does not demonstrate that there are no residual systematics in the map, it does show that any systematic is suppressed far below the white-noise level in a single map voxel. If there were any significant correlated noise in the maps, that would show up as extra variance in this histogram. However, while these results do show that correlated noise is not a major problem, they do not rule out other systematic effects on larger scales in the map.

7. Summary and Conclusions

We have presented the data analysis pipeline used to process the first-season COMAP observations with respect to high-redshift CO emission, from raw TOD to final calibrated maps. This pipeline implements four main steps (calibration, filtering, data selection, and mapmaking), each of which is designed to optimally exploit the unique instrumental capabilities of the COMAP instrument. For instance, calibration is performed using a combination of frequent comparison with a hardware calibrator and real-time total-power measurements. The filtering procedures explicitly exploit the multifeed and multi-frequency design of the COMAP instrument to reject common-mode contaminants, resulting in data that are strongly dominated by uncorrelated white noise after filtering. Finally, both the data selection and mapmaking processes directly use this fact to produce near-optimal goodness-of-fit statistics and pixelized sky maps with high computational efficiency.

By applying this pipeline to data from the first observing season, we have demonstrated a key goal of the Pathfinder: that the noise level integrates down with time as expected for uncorrelated white noise. A careful analysis of the data and observing efficiencies obtained in Season 1 has allowed us to forecast the performance of the Pathfinder taking into account expected and already-implemented improvements to the instrument, analysis, and observing strategy. Based on this forecast and on models for the CO emission at $z \sim 3$, the Pathfinder is expected to achieve a detection of the CO (1–0) auto-power spectrum by the end of the 5 yr observing campaign (Chung et al. 2022).

We thank graduated master student Erik Levén for his contribution to this work. This material is based on work supported by the National Science Foundation under grant Nos. 1517108, 1517288, 1517598, 1518282, and 1910999 and by the Keck Institute for Space Studies under "The First Billion Years: A Technical Development Program for Spectral Line Observations."

Parts of the work were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration, and funded through the internal Research and Technology Development program. D.T.C. is supported by a CITA/Dunlap Institute postdoctoral fellowship. The Dunlap Institute is funded through an endowment established by the David Dunlap family and the University of Toronto. C.D. acknowledges support from an STFC Consolidated Grant (ST/P000649/1).

J.B., H.K.E., M.K.F., H.T.I., J.G.S.L., M.R., N.-O.S., D.W., and I.K.W. acknowledge support from the Research Council of Norway through grants 251328 and 274990 and from the European Research Council (ERC) under the Horizon 2020 Research and Innovation Program (grant agreement No. 819478, COSMOGLOBE). J.G. acknowledges support from the University of Miami and is grateful to Hugh Medrano for assistance with cryostat design. S.H. acknowledges support from an STFC Consolidated Grant (ST/P000649/1). J. Kim is supported by a Robert A. Millikan Fellowship from Caltech. At JPL, we are grateful to Mary Soria for assembly work on the amplifier modules and to Jose Velasco, Ezra Long, and Jim Bowen for the use of their amplifier test facilities. H.P. acknowledges support from the Swiss National Science Foundation through Ambizione grant PZ00P2_179934. P.C. B. is supported by the James Arthur Postdoctoral Fellowship. We thank Isu Ravi for her contributions to the warm electronics and antenna drive characterization. The Scientific color maps `roma` and `tokyo` (Crameri 2021) are used in this study to prevent visual distortion of the data and exclusion of readers with color-vision deficiencies (Crameri et al. 2020).

We also want to thank the anonymous referee, whose comments and suggestions have helped to significantly improve and clarify this manuscript.

Software: Matplotlib (Hunter 2007); Astropy, a community-developed core Python package for astronomy (Astropy Collaboration et al. 2013).

Appendix
System Temperature Distribution

Figure 21 shows a histogram of the sideband-averaged system temperature, T_{sys}, for all Season 1 observations. From the radiometer equation (Equation (9)) we see that the system temperature quantifies the noise level of the observations, so this histogram shows the distribution of the sensitivity of our measurements. Some of these variations are from the different noise levels of the different feeds and sidebands, but most of the variation comes from the different optical depth of the atmosphere when observing at different elevations. Note that the sideband averages were taken after the frequency masking that happens during the low-level filtering. We also exclude observations at elevations higher than 65° and lower than 35°, as these are not used in the final analysis. We see that 95% of the recorded system temperatures are in the range 34–60 K, with a median value of 44 K.

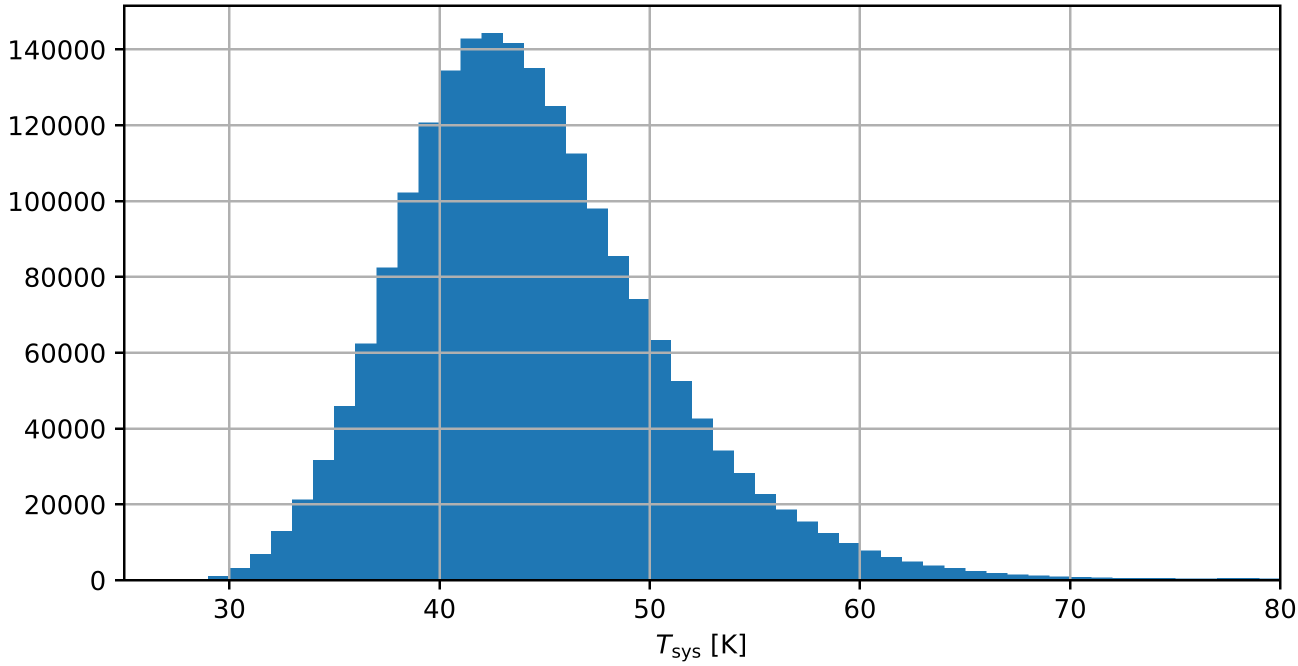


Figure 21. Histogram of sideband-averaged system temperature, T_{sys}, for all Season 1 observations.

ORCID iDs

Marie K. Foss https://orcid.org/0000-0001-8896-3159
Håvard T. Ihle https://orcid.org/0000-0003-3420-7766
Kieran A. Cleary https://orcid.org/0000-0002-8214-8265
Hans Kristian Eriksen https://orcid.org/0000-0003-2332-5281
Stuart E. Harper https://orcid.org/0000-0001-7911-5553
Junhan Kim https://orcid.org/0000-0002-4274-9373
James W. Lamb https://orcid.org/0000-0002-5959-1285
Liju Philip https://orcid.org/0000-0001-7612-2379
Nils-Ole Stutzer https://orcid.org/0000-0001-5301-1377
Bade D. Uzgil https://orcid.org/0000-0001-8526-3464
Duncan J. Watts https://orcid.org/0000-0002-5437-6121
Ingunn K. Wehus https://orcid.org/0000-0003-3821-7275
J. Richard Bond https://orcid.org/0000-0003-2358-9949
Patrick C. Breysse https://orcid.org/0000-0001-8382-5275
Dongwoo T. Chung https://orcid.org/0000-0003-2618-6504
Clive Dickinson https://orcid.org/0000-0002-0045-442X
Delaney A. Dunne https://orcid.org/0000-0002-5223-8315
Andrew I. Harris https://orcid.org/0000-0001-6159-9174
Anthony C. S. Readhead https://orcid.org/0000-0001-9152-961X
Hamsa Padmanabhan https://orcid.org/0000-0002-8800-5740
Timothy J. Pearson https://orcid.org/0000-0001-5213-6231
Thomas J. Rennie https://orcid.org/0000-0002-1667-3897

References

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Battye, R. A., Davies, R. D., & Weller, J. 2004, MNRAS, 355, 1339
Breysse, P. C., Kovetz, E. D., & Kamionkowski, M. 2014, MNRAS, 443, 3506
Chung, D., Li, T., Viero, M. P., Church, S. E., & Wechsler, R. H. 2017, ApJ, 846, 60
Chung, D. T., Breysse, P. C., Cleary, K. A., et al. 2022, ApJ, 933, 186
Chung, D. T., Breysse, P. C., Ihle, H. T., et al. 2021, ApJ, 923, 188
Chung, D. T., Viero, M. P., Church, S. E., et al. 2019, ApJ, 872, 186
Cleary, K. A., Borowska, J., Breysse, P. C., et al. 2022, ApJ, 933, 182
Crameri, F. 2021, Scientific Colour Maps, v.7.0.0, Zenodo, doi:10.5281/zenodo.1243862
Crameri, F., Shephard, G. E., & Heron, P. J. 2020, NatCo, 11, 5444
Gebhardt, K., Mentuch Cooper, E., Ciardullo, R., et al. 2021, ApJ, 923, 217
Hill, G. J., Gebhardt, K., Komatsu, E., et al. 2008, in ASP Conf. Ser. 399, Panoramic Views of Galaxy Formation and Evolution, ed. T. Kodama, T. Yamada, & K. Aoki (San Francisco, CA: ASP), 115
Hill, G. J., Lee, H., MacQueen, P. J., et al. 2021, AJ, 162, 298
Hunter, J. D. 2007, CSE, 9, 90
Ihle, H. T., Borowska, J., Cleary, K. A., et al. 2022, ApJ, 933, 185
Keating, G. K., Bower, G. C., Marrone, D. P., et al. 2015, ApJ, 814, 140
Keating, G. K., Marrone, D. P., Bower, G. C., & Keenan, R. P. 2020, ApJ, 901, 141
Keating, G. K., Marrone, D. P., Bower, G. C., et al. 2016, ApJ, 830, 34
Keenan, R. P., Keating, G. K., & Marrone, D. P. 2022, ApJ, 927, 161
Kovetz, E., Breysse, P. C., Lidz, A., et al. 2019, BAAS, 51, 101
Kovetz, E. D., Viero, M. P., Lidz, A., et al. 2017, arXiv:1709.09066
Lamb, J. W., Cleary, K. A., Woody, D. P., et al. 2022, ApJ, 933, 183
Li, T. Y., Wechsler, R. H., Devaraj, K., & Church, S. E. 2016, ApJ, 817, 169
Lidz, A., Furlanetto, S. R., Oh, S. P., et al. 2011, ApJ, 741, 70
Loeb, A., & Wyithe, J. S. B. 2008, PhRvL, 100, 161301
Madau, P., Meiksin, A., & Rees, M. J. 1997, ApJ, 475, 429
Moradinezhad Dizgah, A., & Keating, G. K. 2019, ApJ, 872, 126
Moradinezhad Dizgah, A., Nikakhtar, F., Keating, G. K., & Castorina, E. 2022, JCAP, 2022, 026
Padmanabhan, H. 2018, MNRAS, 475, 1477
Penzias, A. A., & Burrus, C. A. 1973, ARA&A, 11, 51
Peterson, J. B., Bandura, K., & Pen, U. L. 2006, arXiv:astro-ph/0606104
Pullen, A. R., Chang, T.-C., Doré, O., & Lidz, A. 2013, ApJ, 768, 15
Rasmussen, M. 2020, Master's thesis, University of Oslo
Rennie, T. J., Harper, S. E., Dickinson, C., et al. 2022, ApJ, 933, 187
Riechers, D. A., Pavesi, R., Sharon, C. E., et al. 2019, ApJ, 872, 7
Righi, M., Hernández-Monteagudo, C., & Sunyaev, R. A. 2008, A&A, 489, 489
Silva, M. B., Baumschlager, B., Cleary, K. A., et al. 2021, arXiv:2111.05354
Sun, G., Hensley, B. S., Chang, T.-C., Doré, O., & Serra, P. 2019, ApJ, 887, 142
Visbal, E., & Loeb, A. 2010, JCAP, 11, 016
Woody, D. P., Beasley, A. J., Bolatto, A. D., et al. 2004, Proc. SPIE, 5498, 30
Yang, S., Popping, G., Somerville, R. S., et al. 2022, ApJ, 929, 140

Paper II

COMAP Early Science. IV. Power Spectrum Methodology and Results

The Astrophysical Journal, 933:185 (14pp), 2022 July 10 https://doi.org/10.3847/1538-4357/ac63c5

COMAP Early Science. IV. Power Spectrum Methodology and Results

Håvard T. Ihle[1], Jowita Borowska[1], Kieran A. Cleary[2], Hans Kristian Eriksen[1], Marie K. Foss[1], Stuart E. Harper[3], Junhan Kim[2], Jonas G. S. Lunde[1], Liju Philip[4], Maren Rasmussen[1], Nils-Ole Stutzer[1], Bade D. Uzgil[5,6], Duncan J. Watts[1], Ingunn Kathrine Wehus[1], J. Richard Bond[7,8], Patrick C. Breysse[9], Morgan Catha[10], Sarah E. Church[11], Dongwoo T. Chung[7], Clive Dickinson[3], Delaney A. Dunne[2], Todd Gaier[4], Joshua Ott Gundersen[12], Andrew I. Harris[13], Richard Hobbs[10], James W. Lamb[10], Charles R. Lawrence[4], Norman Murray[7], Anthony C. S. Readhead[2], Hamsa Padmanabhan[14], Timothy J. Pearson[2], Thomas J. Rennie[3], and David P. Woody[10]
(COMAP Collaboration)

[1] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway; h.t.ihle@astro.uio.no
[2] California Institute of Technology, Pasadena, CA 91125, USA
[3] Jodrell Bank Centre for Astrophysics, Alan Turing Building, Department of Physics and Astronomy, School of Natural Sciences, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[4] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[5] California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
[6] National Radio Astronomy Observatory, Pete V. Domenici Array Science Center, P.O. Box 0, Socorro, NM 87801, USA
[7] Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada
[8] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[9] Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA
[10] Owens Valley Radio Observatory, California Institute of Technology, Big Pine, CA 93513, USA
[11] Kavli Institute for Particle Astrophysics and Cosmology & Physics Department, Stanford University, Stanford, CA 94305, USA
[12] Department of Physics, University of Miami, 1320 Campo Sano Avenue, Coral Gables, FL 33146, USA
[13] Department of Astronomy, University of Maryland, College Park, MD 20742, USA
[14] Departement de Physique Théorique, Université de Genève, 24 Quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
Received 2021 November 19; revised 2022 February 25; accepted 2022 March 25; published 2022 July 13

Abstract

We present the power spectrum methodology used for the first-season COMAP analysis, and assess the quality of the current data set. The main results are derived through the Feed–Feed Pseudo-Cross-Spectrum (FPXS) method, which is a robust estimator with respect to both noise modeling errors and experimental systematics. We use effective transfer functions to take into account the effects of instrumental beam smoothing and various filter operations applied during the low-level data processing. The power spectra estimated in this way have allowed us to identify a systematic error associated with one of our two scanning strategies, believed to be due to residual ground or atmospheric contamination. We omit these data from our analysis and no longer use this scanning technique for observations. We present the power spectra from our first season of observing, and demonstrate that the uncertainties are integrating as expected for uncorrelated noise, with any residual systematics suppressed to a level below the noise. Using the FPXS method, and combining data on scales $k = 0.051–0.62\ \mathrm{Mpc}^{-1}$, we estimate $P_{\mathrm{CO}}(k) = -2.7 \pm 1.7 \times 10^4\ \mu\mathrm{K}^2\ \mathrm{Mpc}^3$, the first direct 3D constraint on the clustering component of the CO(1–0) power spectrum in the literature.

Unified Astronomy Thesaurus concepts: CO line emission (262); Cosmological evolution (336); High-redshift galaxies (734); Molecular gas (1073); Radio astronomy (1338)

1. Introduction

Intensity mapping aims to map out large 3D volumes using bright emission lines as tracers of large-scale matter distribution (Madau et al. 1997; Battye et al. 2004; Peterson et al. 2006; Loeb & Wyithe 2008; Kovetz et al. 2017, 2019). One promising set of lines comprises the rotational transitions of the carbon monoxide (CO) molecule. CO traces cold molecular gas, and is closely linked to star formation (Carilli & Walter 2013).

The CO Mapping Array Project (COMAP) is an intensity mapping experiment targeting CO. This paper, one of a set associated with the first-season COMAP analysis, presents the methodology used to constrain the CO power spectrum with COMAP data. An overview of the COMAP experiment is presented by Cleary et al. (2022), while the COMAP instrument is described by Lamb et al. (2022).

The low-level COMAP data-processing pipeline is summarized by Foss et al. (2022). This pipeline converts raw uncalibrated observations into 3D maps, using redshifted CO line emission from distant galaxies as a tracer of the cosmic density field. Since the first-season COMAP instrument observes at frequencies between 26 and 34 GHz, and the rotational CO(1–0) transition has a rest frequency of 115 GHz, the current measurements trace galaxy formation at redshifts between $z = 2.4$ and 3.4, during "the epoch of galaxy assembly." The current limits, forecasts, and modeling at these redshifts is discussed in Chung et al. (2022), while a future phase of COMAP, targeting "the epoch of reionization," is discussed in Breysse et al. (2022). The use of this instrument for a galactic survey is presented in Rennie et al. (2022).

One common and powerful quantity used to characterize the statistical properties of such 3D cosmic maps is the so-called power spectrum (or the two-point correlation function), which measures the strength of fluctuations as a function of physical distance (e.g., Lidz et al. 2011; Pullen et al. 2013; Li et al. 2016; Bernal et al. 2019; Chung 2019; Ihle et al. 2019; Uzgil et al. 2019; Gong et al. 2020; Keenan et al. 2022; Yang et al. 2021; Moradinezhad Dizgah et al. 2022). For an isotropic and Gaussian random field, this function quantifies all statistically relevant information in the original data set, but with a far smaller number of data points, and it thus represents a dramatic compression of the full data set. For non-Gaussian fields, additional information can be extracted by the use of other statistics (Breysse et al. 2017, 2019; Ihle et al. 2019; Sato-Polito & Bernal 2022). Even for non-Gaussian fields, however, such as the galactic density field, the power spectrum encapsulates a large fraction of the important information, and it is therefore an efficient tool even for such fields.

However, while compressing hundreds of terabytes of raw data into a handful of power spectrum coefficients certainly makes the interpretation of the data easier in terms of theoretical comparisons, it also makes the final estimates highly sensitive to small systematic effects and instrumental noise. To guide our intuition, we note that current theories predict an intrinsic CO standard deviation per resolution element of no more than a few microkelvin (Breysse et al. 2014; Li et al. 2016; Chung et al. 2022), which is to be compared with a typical system temperature of 44 K for the COMAP instrument; or atmospheric fluctuations of a few kelvin; or sidelobe contributions of a few millikelvin. All such effects must therefore be suppressed by many orders of magnitude in order to establish robust astrophysical constraints.

As described by Lamb et al. (2022) and Foss et al. (2022), the COMAP focal plane consists of 19 different feed horns, arranged in a hexagonal pattern, with about 12′ sky separation between the closest feeds. The signal entering each feed horn is sent through its own signal chain, with its own amplifiers and digital high-resolution spectrometers. Each such signal chain is typically referred to as a "feed." As such, the instrument has many unique features that makes it suited to this process. A few important examples include the highly efficient spectroscopic rejection of common-mode signals, several semi-independent feeds, a configurable scanning strategy, and frequent usage of hardware calibrators. Still, the rejection of systematic errors at the microkelvin level is highly challenging, and the current paper describes several algorithmic methods that can be applied to improve the robustness of the final results.

The rest of the paper is organized as follows. In Section 2, we review various aspects of power spectrum methods and present our adopted COMAP power spectrum estimator, the Feed–Feed Pseudo-Cross-Spectrum (FPXS). Power spectra estimated using data from COMAP's first observing season are presented in Section 3, and we conclude in Section 4.

2. Methods

We begin our discussion with an overview of the fundamental algorithms used for COMAP power spectrum estimation. For other recent examples of the use of power spectrum analysis on intensity mapping data, see, e.g., Mertens et al. (2020), Keating et al. (2020), and Keenan et al. (2022).

2.1. Autospectrum Analysis

Let m_{ijk} denote a 3D map, and let us call each resolution element in this map a voxel. i, j, and k are then the voxel indices. We define the power spectrum $P(\boldsymbol{k})$ of this map to be the variance of its Fourier components, $f_{\boldsymbol{k}}$:

$$P(\boldsymbol{k}) = \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}}}\langle |f_{\boldsymbol{k}}|^2\rangle, \tag{1}$$

where $\boldsymbol{k}$ is the wavevector of a given Fourier mode, V_{vox} is the volume of each voxel, and N_{vox} is the total number of voxels.

If we assume that the map is statistically isotropic, then the power spectrum will only be a function of the magnitude of the wavevector, $P(\boldsymbol{k}) = P(k)$. In observational cosmology, we often want to distinguish the angular directions (denoted by the x and y coordinates) from the line-of-sight (LOS) direction (denoted by the z coordinate). This is because the map typically has different properties in different directions; for example, due to instrumental beam effects or redshift space distortions (Hamilton 1998; Chung 2019). It is therefore often useful to define the power spectrum in terms of parallel (LOS) modes, $k_\parallel \equiv |k_z|$, and perpendicular (angular) modes, $k_\perp \equiv \sqrt{k_x^2 + k_y^2}$. We can estimate the power spectrum in a given set of $\boldsymbol{k}$-bins, $\{\boldsymbol{k}_i\}$, from a given map as

$$P(\boldsymbol{k}_i) \approx \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}}N_{\mathrm{modes}}}\sum_{j=1}^{N_{\mathrm{modes}}}|f_{\mathbf{k}_j}|^2 \equiv P_{\boldsymbol{k}_i}, \tag{2}$$

where $P_{\boldsymbol{k}_i}$ is the estimated power spectrum in bin number i and $N_{\mathrm{modes,i}}$ is the number of Fourier components with wavenumber $\boldsymbol{k}_j \approx \boldsymbol{k}_i$ (i.e., in the bin corresponding to wavenumber $\boldsymbol{k}_i$).

Assuming that foreground and systematic contributions have already been removed to negligible levels through preprocessing, the power spectrum of a cleaned line intensity map is typically modeled as a sum of a signal and noise component (assumed to be statistically independent):

$$P(\boldsymbol{k}) = P_{\mathrm{signal}}(\boldsymbol{k}) + P_{\mathrm{noise}}(\boldsymbol{k}). \tag{3}$$

If we are able to estimate the noise power spectrum through independent means—for example, by using a noise model or simulations—we can extract the signal power spectrum simply by subtracting the estimated noise,

$$P_{\mathrm{signal}}(\boldsymbol{k}_i) \approx P_{\boldsymbol{k}_i} - P_{\mathrm{noise}}^{\mathrm{est}}(\boldsymbol{k}_i), \tag{4}$$

where $P_{\mathrm{noise}}^{\mathrm{est}}(\boldsymbol{k}_i)$ is the estimated noise power spectrum in bin number i.

If the map consists of uniformly distributed white noise, then the noise power spectrum is independent of $\boldsymbol{k}$ and given by

$$P_{\mathrm{noise}} = V_{\mathrm{vox}}\sigma_T^2, \tag{5}$$

where σ_T is the white noise standard deviation in each voxel (in units of kelvin). In our case, this magnitude of the white noise level is determined by the radiometer equation,

$$\sigma_T = \frac{T_{\mathrm{sys}}}{\sqrt{\delta_\nu \tau}}, \tag{6}$$

where T_{sys} is the system temperature of the detector; δ_ν is the frequency resolution of each voxel; and τ is the total time each pixel is observed.

In addition to this instrumental noise contribution, there is an intrinsic uncertainty when estimating the signal power spectrum from a map, called sample variance, which arises from the limited number of Fourier modes in the map. Together, these contributions give us the uncertainty of the power spectrum:

$$\sigma_P \equiv \sqrt{\langle (P_{\boldsymbol{k}_i} - P(\boldsymbol{k}_i))^2 \rangle} \approx \underbrace{\frac{P_{\rm noise}(\boldsymbol{k}_i)}{\sqrt{N_{\rm modes}}}}_{\text{Thermal noise}} + \underbrace{\frac{P_{\rm signal}(\boldsymbol{k}_i)}{\sqrt{N_{\rm modes}}}}_{\text{Sample variance}}, \tag{7}$$

where $N_{\rm modes}$ is the number of Fourier modes in bin number i, and the last approximation is exact when the Fourier modes are assumed to be independent Gaussian variables.

If the power spectrum is noise dominated, we can reduce this intrinsic uncertainty in two ways. First, we can observe the same area of sky for a longer time, thus decreasing the noise power spectrum contribution to the uncertainty. Alternatively, we can cover a larger sky area, and thus increase the number of measured Fourier modes. As long as we are noise dominated, a simple analysis suggests that observing a small area for a long time is more efficient for making a first detection than spreading the observations over a larger area. In a realistic situation, however, there are several other factors that must be taken into account, including the choice of angular resolution and scanning strategy constraints, and these will typically limit how small a field it is possible to observe.

Another source of uncertainty in estimating the signal is the accuracy of the estimated noise power spectrum model. If this model is biased or uncertain, then the associated residuals will propagate directly into the estimate of $P_{\rm signal}(\boldsymbol{k}_i)$.

2.2. Pseudospectrum Analysis

As described above, there are several challenges with an autospectrum analysis, as will be discussed both in this and the following sections. First of all, if the noise in the map is not uniform, which it generally is not, the noise power spectrum will be dominated by the parts of the map with the highest noise levels. In order to address this, it is necessary to devise a method that puts more weight on the parts of the map with low noise, and less weight on the parts of the map with high noise.

The standard method of accounting for this is through inverse noise-variance weights. That is, we weigh the map, m, by the noise level map, σ_m (the map given by the expected standard deviation of the white noise in each voxel), before we compute the power spectrum,

$$\tilde{P}_{\boldsymbol{k}_i} = \frac{V_{\rm vox}}{N_{\rm vox} N_{\rm modes}} \sum_{j=1}^{N_{\rm modes}} |\tilde{f}_{\mathbf{k}_j}|^2, \tag{8}$$

where $\tilde{P}$ denotes the pseudospectrum and $\tilde{f}$ are the Fourier components of the noise-weighted map,

$$\tilde{m} \equiv wm \tag{9}$$

and

$$w \equiv \mathcal{N} \frac{1}{\sigma_m^2}. \tag{10}$$

$\mathcal{N}$ is a single overall normalization constant (which we will get back to), and σ_m is, as usual, the noise level map.

On a general note, the term "pseudospectrum" typically refers to a power spectrum estimator that is computed from a biased estimator of the true sky map, and as such is itself biased; see Hivon et al. (2002). This may be contrasted with more conventional power spectrum estimators that aim to estimate the power spectrum of the true sky signal. The statistical information content of the pseudospectrum and the unbiased power spectrum is identical, and the main difference between the two classes of estimators concerns their ease of interpretation; while the unbiased power spectrum may be directly compared with theoretical models and other literature results, the pseudospectrum is experiment dependent, and typically requires simulations for proper statistical interpretation.

In our setting, we use the pseudospectrum to take into account both masked voxels (by setting $\sigma_m \to \infty$ for voxels that are excluded from further analysis) and varying noise levels across the map. Both these operations lead to *mode mixing*, i.e., different-signal Fourier modes are mixed together, and the estimated signal pseudospectrum is therefore a distorted version of the true signal power spectrum. However, since we know exactly how the signal map has been distorted, we can, at least in principle, calculate the exact mode-mixing matrix that is needed to reconstruct the mode mixing and obtain an unbiased signal spectrum from the pseudospectrum (Hivon et al. 2002). How feasible this is for a specific case depends on the details of the map dimensions and computational resources. For more details on mode mixing, see Appendix D.

Although mode mixing does complicate the physical interpretation of the pseudospectrum, there are several ways of dealing with this without having to calculate and invert the full mode-mixing matrix. First of all, if the analysis involves comparisons with signal simulations, then one may simply apply the same weight matrix to each simulation, making the observed and simulated power spectra statistically compatible. Second, if the level of mode mixing is modest, then the pseudospectrum may be an adequate estimator for the signal power spectrum for a given application, especially on smaller scales. This typically holds particularly well for noise-dominated applications, for which a single power estimate covering a large range in k is desired; in that case, the mode mixing often has a minimal effect on the estimates, and the pseudospectrum is often a perfectly valid estimate in its own right. The accuracy of this approximation must be assessed for each use case.

In cases for which the pseudospectrum is intended to be used as a direct estimator, it is necessary to choose a value for the normalization factor $\mathcal{N}$ in Equation (10). Establishing the *formally correct* value for this normalization is not entirely well defined, as you are essentially trying to approximate the effect of an entire matrix with a single number (see Appendix D for more details). However, we can make a simple and fairly reasonable choice as follows:

$$\mathcal{N} = \frac{1}{\sqrt{\left\langle \frac{1}{\sigma_m^4} \right\rangle}}, \tag{11}$$

where $\langle\rangle$ denotes the average over the whole map. To make the results easier to interpret, we therefore apply this normalization to all the results shown in this paper. For analyses that employ the full mode-mixing matrix, or in which the pseudospectrum is compared directly to simulations, this normalization is completely irrelevant.

To roughly estimate the expected level of mode mixing, we calculate the ratio of the pseudospectrum and the autospectrum

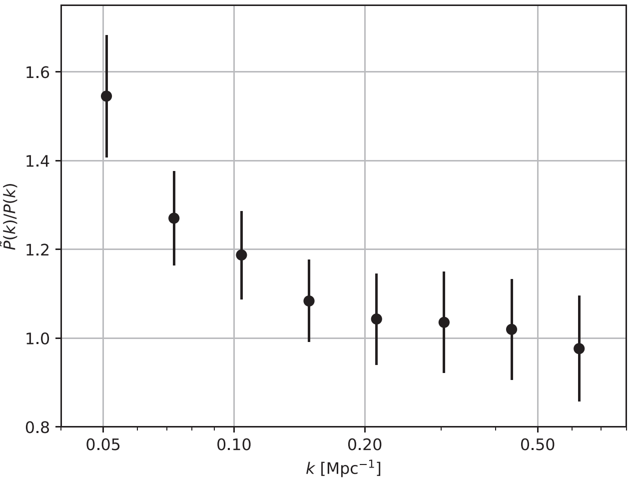


Figure 1. Ratio of the signal pseudospectrum to the signal autospectrum, based on 10 signal realizations.

for 10 signal realization maps. Figure 1 shows the mean and standard deviation of the mode mixing in each of the main power spectrum bins. Overall, we see that at the scales where we have most of our sensitivity, the effect of mode mixing is fairly modest, typically in the 5%–30% range. Thus, even at face value, the pseudospectrum does provide a reasonable order-of-magnitude estimate of the true power spectrum, even if it may not be appropriate for precision analysis. We also note that these results suggest that, if anything, an upper limit obtained by interpreting the pseudospectrum at face value will be a slightly weaker (i.e., more conservative) upper limit than we would get by accounting for the mode mixing.

We leave it for future work to estimate the mode-mixing matrix and undo the mode-mixing bias in the pseudospectra. For the rest of this paper, we will interpret the pseudospectra at face value.

2.3. Cross-spectrum Analysis

A general challenge when using either the auto- or pseudospectrum is that highly accurate estimates of the noise contribution are required to estimate the signal power spectrum. In many cases, this can be very challenging, and any systematic error will directly bias the final signal estimate.

One way to avoid this complication is to use the so-called cross-spectrum, $C(\boldsymbol{k})$. While the power spectrum quantifies the variance of the Fourier components of a single map, the cross-spectrum quantifies the covariance between the Fourier modes of two different maps:

$$C(\boldsymbol{k}) = \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}}} \langle \mathrm{Re}\{f_{1\boldsymbol{k}}^{*} f_{2\boldsymbol{k}}\} \rangle \approx \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}} N_{\mathrm{modes}}} \sum_{j=1}^{N_{\mathrm{modes}}} \mathrm{Re}\{f_{1\boldsymbol{k}_j}^{*} f_{2\boldsymbol{k}_j}\} \equiv C_{\boldsymbol{k}_i}. \tag{12}$$

Here, Re{ } denotes the real part of a complex number, and f_1 and f_2 are the Fourier components of two maps, m_1 and m_2.

Clearly, if m_1 and m_2 are identical, then the cross-spectrum is equivalent to the autospectrum. The advantage of the cross-spectrum, however, is that if the maps m_1 and m_2 are made from different data, the noise contributions are independent, and they do not contribute to the mean of the cross-spectrum, only to its variance. Therefore, it is not necessary to estimate and subtract the noise power spectrum to obtain an unbiased signal estimate, but rather

$$\langle C_{\boldsymbol{k}_i} \rangle = P_{\mathrm{signal}}(\boldsymbol{k}_i). \tag{13}$$

Of course, a proper noise estimate is still necessary for uncertainty estimation, but the requirements on this are typically far less stringent than for the estimator mean.

Although the cross-spectrum significantly reduces the precision needed when estimating the noise power spectrum, we do pay a price in the form of somewhat lower intrinsic sensitivity. For instance, when splitting the data into two independent parts, and cross-correlating these, we do lose a factor of at least $\sqrt{2}$ from the fact that we do not exploit the autocorrelations within each data set separately. Fortunately, this problem can be remedied by splitting the data into more independent maps, and averaging the cross-spectra of all possible combinations. A lower limit on the cross-spectrum sensitivity is given by

$$\sigma_C^{N_{\mathrm{split}}} \geqslant \sqrt{\frac{1}{1 - 1/N_{\mathrm{split}}}}\, \sigma_P, \tag{14}$$

where N_{split} is the number of different map splits and σ_P is the optimal sensitivity of the autospectrum derived from the full data set.

The cross-spectrum has some other very important advantages with respect to the autospectrum, as well. As discussed in the introduction, one of the major challenges for an experiment like COMAP, in which we have to integrate down the noise by several orders of magnitude in order to measure a small signal, are systematic errors. However, since the cross-spectrum may only be biased by structures common to the two maps, one can try to ensure that any known systematic effects contribute to the two maps independently. In that case, the systematic effects will not bias the signal estimate. Combining this insight with splitting the data into multiple parts allows us to design a power spectrum statistic that is far more robust to systematics than the autospectrum.

We define a pseudo-cross-spectrum in an analogous manner to the pseudo-auto-spectrum. The only subtlety is that we make sure to apply the same weight map, w, for both maps. Explicitly, we adopt the following weight map,

$$w_{1,2} \propto \frac{1}{\sigma_{m_1} \sigma_{m_2}}, \tag{15}$$

for both m_1 and m_2 when calculating the pseudo-cross-spectrum, $\tilde{C}_{\boldsymbol{k}_i}$.

2.4. The Feed–Feed Pseudo-Cross-Spectrum Method

The idea of the FPXS method is to combine all the insights from the preceding sections to construct a single statistic for the CO signal, which has a high intrinsic sensitivity, uses proper noise weighting, and is robust against instrumental and other systematic errors. In that respect, we first note that the COMAP focal plane consists of 19 feeds, each with its own amplifiers and detectors. Furthermore, many systematic errors are particular to each feed, due to different passbands, amplifiers, cables, beams, etc. We may therefore split the data according to the feeds (i.e., make one map per feed), and then compute the cross-spectra of all the different feed combinations, without ever correlating two maps from the same feed.

Second, we also note that one of the most troublesome systematic errors for COMAP is ground pickup. This is mainly because the ground contamination correlates with the pointing, and it therefore does not average down in the same way as any systematic error that is random in the time domain (and hence independent in different observations). We can make ourselves as robust as possible against any residual ground signal in our map by also splitting the data by the elevation of the observations, so that we never take the cross-spectrum of two different data sets taken at the same elevation.[15]

With these considerations in mind, we define the following procedure for calculating the FPXS:

1. We split the data into disjoint sets, sorted according to elevation. For simplicity, we assume for now that we split the data into two sets, A and B, where A contains all the observations taken at elevations below the median elevation, and B contains all the observations from the higher elevations. We can easily generalize this to a case where we split the data into more than two sets.
2. For each set, A and B, we generate maps for each of the 19 feeds. We denote the different maps according to data set and feed, such that A_{13} indicates the map that combines all data from data set A for feed number 13.
3. We then calculate the pseudo-cross-spectrum, $\tilde{C}^{ij}_{\boldsymbol{k}_i}$, for all the different map combinations of A_i and B_j where $i \neq j$.
4. Next, we compute the average pseudo-cross-spectrum, $\tilde{C}_{\boldsymbol{k}_i}$, by noise weighting all the different cross-spectra:

$$\tilde{C}^{\mathrm{FPXS}}_{\boldsymbol{k}_i} = \left(\sum_{i\neq j} \frac{1}{\sigma^2_{\tilde{C}^{ij}_{\boldsymbol{k}_i}}}\right)^{-1} \sum_{i\neq j} \frac{\tilde{C}^{ij}_{\boldsymbol{k}_i}}{\sigma^2_{\tilde{C}^{ij}_{\boldsymbol{k}_i}}}. \tag{16}$$

Here, $\sigma_{\tilde{C}^{ij}_{\boldsymbol{k}_i}}$ is the uncertainty (standard deviation) in $\boldsymbol{k}$-bin number i of the pseudo-cross-spectrum of the maps A_i and B_j, and the sum is over all combinations of i and j except the cases where $i = j$. Under the naive assumption that all cross-spectra are independent, the uncertainty of the combined cross-spectrum is given by

$$\sigma_{\tilde{C}^{\mathrm{FPXS}}_{\boldsymbol{k}_i}} = \left(\sum_{i\neq j} \frac{1}{\sigma^2_{\tilde{C}^{ij}_{\boldsymbol{k}_i}}}\right)^{-1/2}. \tag{17}$$

The data can of course be split in other ways, to make ourselves less susceptible to other systematic effects, but we have found that using the feeds and elevation splits yields good results for the current data set.

2.5. White Noise Simulations

Until now, we have not discussed how to estimate the noise power spectrum and the corresponding noise uncertainty of the power spectrum. In general, estimating the noise power spectrum precisely is very difficult, since one needs to take into account not only the intrinsic white noise level of the data, but also the effect of the different filtering procedures in the low-level data analysis, as well as any correlated noise contribution.

Since we use a cross-spectrum method, however, the noise spectrum is only used to estimate the uncertainties of the power spectrum, not its mean level, and the requirements on the absolute noise spectrum are therefore somewhat relaxed. Explicitly, if we make an error of a few percent in our noise estimate, we will not bias the estimated signal spectrum, only misestimate the error bars by a few percent. While clearly not ideal, this is usually not critical, considering all the other simplifying assumptions introduced in the analysis. On the other hand, if we had adopted an autospectrum method, an error of a few percent on the noise power spectrum could easily have rendered our signal estimate unusable, even in the case of very high intrinsic sensitivity.

For this reason, we therefore adopt a simple approach to noise power spectrum estimation: we assume that the noise in the maps is uncorrelated white noise, and generate noise simulations, m_i, by drawing random samples in each voxel from a Gaussian distribution with zero mean and a standard deviation given by the value of the noise level map, σ_m. We then estimate error bars by generating a large number of noise simulation maps, calculating the power spectrum from each, and finally taking the standard deviation in each $\boldsymbol{k}$-bin of interest. This gives us uncertainties on the noise contribution to each power spectrum bin, but neglects the intrinsic uncertainty in the signal power spectrum itself. However, as we are still completely noise dominated, this intrinsic uncertainty of the signal spectrum should be negligible.

2.6. Transfer Functions

Until now, we have assumed that the sky maps produced by the low-level analysis pipeline are unbiased. For multiple reasons, this is not the case. First of all, the instrument does not have infinite resolution, and the instrumental beam will therefore smooth out the signal on small angular scales. The same effect occurs due to the finite spectral resolution of the instrument in the frequency dimension. Second, the various filters and mapmaking procedures in the analysis pipeline generally remove some of the signal, mostly on larger angular and spectral scales. In the following, we take these effects into account through so-called transfer functions. These are functions in the $k_\parallel$–$k_\perp$ plane that quantify the fraction of the signal power that is retained in each $\boldsymbol{k}$-bin, and allow us to establish unbiased estimates of the power spectrum from biased sky maps.

In general, a transfer function, $T(\boldsymbol{k})$, is defined through the following relation:

$$\langle P_{\boldsymbol{k}} \rangle = T(\boldsymbol{k}) P_{\mathrm{signal}}(\boldsymbol{k}) + P_{\mathrm{noise}}(\boldsymbol{k}), \tag{18}$$

where $P_{\boldsymbol{k}}$ is the power spectrum calculated from the final map and $P_{\mathrm{signal}}(\boldsymbol{k})$ is the actual physical signal power spectrum. We decompose the full transfer function into different parts, and derive each separately. We then multiply the transfer functions together to get the full transfer function.

In writing down Equation (18) with a transfer function, $T(\boldsymbol{k})$, that is not a function of the signal, we have implicitly assumed that the effects we are accounting for using the transfer function are linear, that they do not depend on the properties of the signal, and that they can therefore be estimated using any signal model. While this is a good approximation in many cases (e.g., the beam effect is purely linear and most of the low-level filters are linear, assuming the same scanning pattern), it is only an approximation. However, even any residual theoretical dependence

[15] The ground contamination also depends on azimuth, but since most of the problematic ground contamination happens at the highest or lowest elevations, it is most natural to divide the data according to elevation.

on the input signal will typically be small in practice, since the signal power spectrum in any reasonable model is very smooth and has no sharp features. The most important effects to get right when estimating the transfer function are the noise distribution and scanning pattern. This is even more important because we are working with pseudospectra, where the noise level will affect the weighting and the mode mixing. That is why, even though it is costly to produce simulated data, we use about 63 hr of simulated data (thus ensuring a realistic scanning pattern as well as noise distribution) when we estimate the pipeline transfer function. Since we are using the pseudospectrum anyway, and not accounting for the mode mixing, we are already accepting errors of the order of 10%, which puts less stringent constraints on the precision of the rest of our procedures.

2.6.1. Instrumental Beam Transfer Function

Due to the finite resolution of the instrument, we cannot measure the cosmological signal on the smallest angular scales. In order to take this effect into account, we introduce a beam transfer function. For now, we assume the beam to be both achromatic (i.e., constant in frequency) and azimuthally symmetric. We construct an azimuthally symmetric beam model by averaging the full 2D (azimuth-elevation) beam model (Lamb et al. 2022) and inserting an exponential cutoff at around 30′.

The ambient load calibration discussed in Foss et al. (2022) measures all of the power entering the feed horns, including the power that comes from the ground and all of the sky above the horizon. However, any power on scales larger than the modes we are sensitive to is essentially lost. To get a proper, scale-dependent, calibration, our beam model is normalized using observations of Jupiter and TauA, which show that 72% of the power is in the central 6′.4 of the beam (Rennie et al. 2022).

In addition, by including the beam model out to about 30′, we take into account the roughly 10% of extra power that is retained at larger angular scales. We could include the beam model further out, but we are already hitting diminishing returns, so not much more would be gained.

Our (unnormalized) beam model can be seen in the top panel of Figure 2. The corresponding beam transfer function is estimated using signal-only simulations. That is, we generate a large number of 3D signal realizations and convolve our azimuthally symmetric beam model with the angular dimensions of the map. We then calculate the power spectrum of each of the signal realization maps, with and without beam smoothing. The estimated transfer function is given as the ratio of the average of these:

$$T^{\rm beam}(\boldsymbol{k}) \approx \frac{\langle P_{\boldsymbol{k}}^{\rm signal,beam}\rangle}{\langle P_{\boldsymbol{k}}^{\rm signal}\rangle}, \tag{19}$$

where $P_{\boldsymbol{k}}^{\rm signal,beam}$ is the power spectrum calculated from a beam-smoothed signal realization map and $P_{\boldsymbol{k}}^{\rm signal}$ is the power spectrum calculated from the non-smoothed one.

Figure 2 shows the beam transfer function derived using 100 signal simulations. We see that the beam smoothing suppresses the power on small angular scales, corresponding to the main beam FWHM of about 4′.5. We also see that although we lose sensitivity from our main beam efficiency, we retain some of this power on larger scales by making use of a beam model up to around 30′.

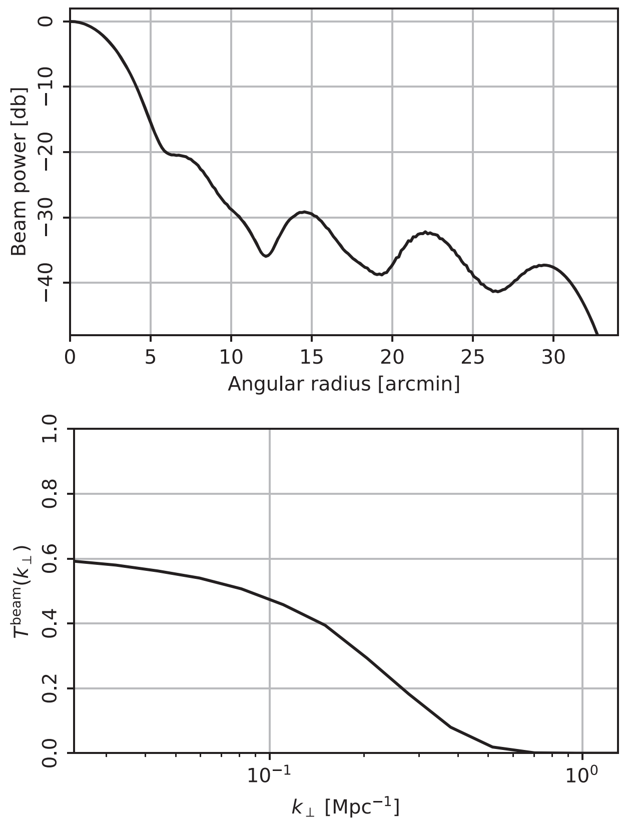


Figure 2. Top: radially symmetric instrumental beam model. Bottom: the resulting beam transfer function, after taking into account the main beam efficiency.

2.6.2. Frequency Resolution Transfer Function

Our current analysis, for simplicity and computational efficiency, uses fairly wide bins in frequency, of 31.25 MHz. This can be compared to intrinsic CO line widths of order 30 MHz (Chung et al. 2021), which will give the smallest scales present in the CO signal we are trying to observe. Once we are ready to claim a detection, we will increase the frequency resolution by at least a factor of 2 to get slightly more sensitivity to the small-scale CO signal, but for now this is not a high priority. We will, however, take into account the bias induced by the current bin size. Often such effects are taken into account by applying an analytic pixel window function, but this is not sufficient here, since the presence of structure on scales smaller than our bins means that some of this power can be aliased into our power spectrum. As the effect depends on the small-scale structure of the signal, there is no model-independent way to take this effect into account, and we will have to use simulations.

We estimate the frequency binning transfer function, $T^{\rm freq}(\boldsymbol{k})$, by comparing power spectra of the simulated signal (using the default model in Chung et al. 2022) on a high-resolution frequency grid to power spectra of the simulated signal on our current frequency grid, both binned in our current $k_\parallel \times k_\perp$ bins. The transfer function derived using 50 such signal simulations is shown in Figure 3. We see a decrease in power toward smaller-LOS scales, but with an increase in the final bin, which we believe is the effect of aliasing the smaller-scale structure into this bin.

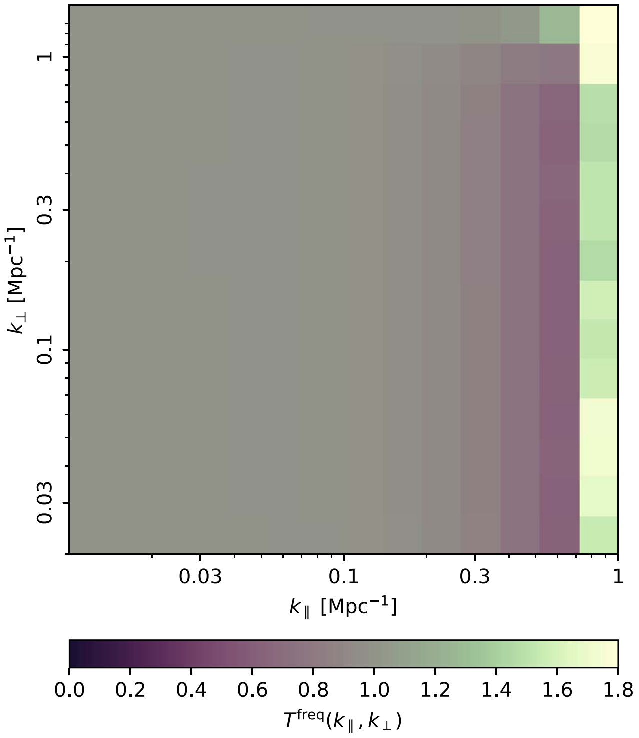


Figure 3. Frequency binning transfer function.

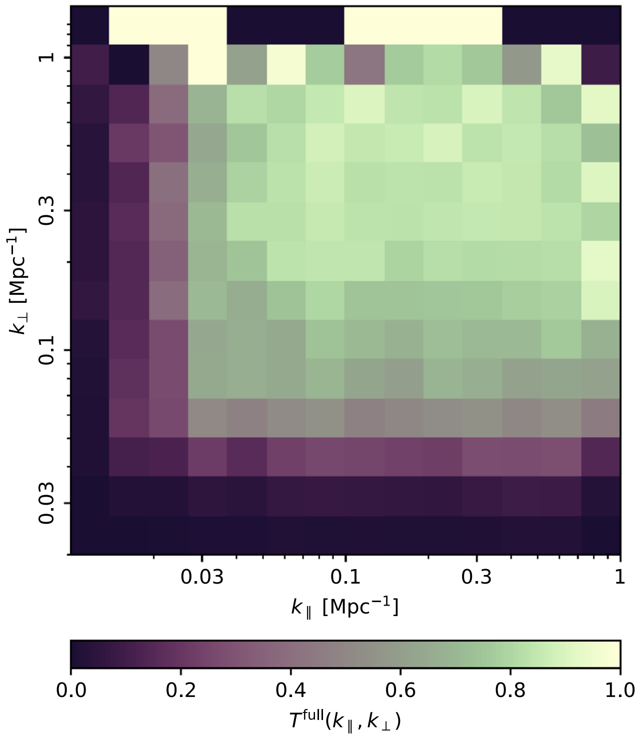


Figure 4. Pipeline transfer function for the cylindrically averaged power spectrum for constant-elevation scans. This transfer function is based on a single signal realization and roughly 3 hr of data.

2.6.3. Pipeline Transfer Function

Each step of the analysis pipeline—including low-level filtering, calibration, and mapmaking—affects how much of the true sky signal is present in the final maps and power spectra. We estimate the transfer function of these operations by processing the sum of the raw data and a known signal-only time-ordered simulation through the analysis pipeline, following the exact same procedure as for the raw data alone. The pipeline transfer function may then be estimated as

$$T^{\rm pipeline}(\boldsymbol{k}) \approx \left\langle \frac{P_{\boldsymbol{k}}^{\rm full} - P_{\boldsymbol{k}}^{\rm noise}}{P_{\boldsymbol{k}}^{\rm signal}} \right\rangle, \tag{20}$$

where $P_{\boldsymbol{k}}^{\rm full}$ is the power spectrum calculated from the maps derived from the raw data with added signal, $P_{\boldsymbol{k}}^{\rm noise}$ is the power spectrum derived from the same data but without the added signal, and $P_{\boldsymbol{k}}^{\rm signal}$ is the power spectrum derived from the raw signal simulation that was added to the raw data.

In Figure 4, the 2D binned pipeline transfer function for the Coudé Echelle Spectrometer (CES) data is shown. The transfer function peaks at intermediate ks, with efficiencies of ∼0.8–0.85 around the peak region. We see that we lose the largest scales both in the angular and LOS directions. This is due to the various filters applied to the time-ordered data to remove correlated noise and systematics, in addition to the effects of the scanning strategy. For more details, see Foss et al. (2022).

2.6.4. Unbiased Signal Estimate

Figure 5 shows the full transfer function, combining all the effects discussed above. Correcting the FPXS with the above transfer function, we can establish an unbiased estimate of the signal pseudospectrum,

$$\tilde{P}_{\rm signal}(\boldsymbol{k}) \approx \tilde{C}_{\boldsymbol{k}} \equiv \frac{\tilde{C}_{\boldsymbol{k}}^{\rm FPXS}}{\tilde{T}_{\boldsymbol{k}}^{\rm full}}, \tag{21}$$

where $\tilde{P}_{\rm signal}(\boldsymbol{k})$ is the signal pseudospectrum and $\tilde{T}_{\boldsymbol{k}}^{\rm full} = \tilde{T}_{\boldsymbol{k}}^{\rm beam}\tilde{T}_{\boldsymbol{k}}^{\rm freq}\tilde{T}_{\boldsymbol{k}}^{\rm pipeline}$ is the full estimated transfer function for the pseudospectrum. The uncertainty of this signal estimate is given by

$$\sigma_{\tilde{P}_{\rm signal}(\boldsymbol{k})} = \sigma_{\tilde{C}_{\boldsymbol{k}}} \equiv \frac{\sigma_{\tilde{C}_{\boldsymbol{k}}^{\rm FPXS}}}{\tilde{T}_{\boldsymbol{k}}^{\rm full}}. \tag{22}$$

2.6.5. Spherical Averaging

Due to the transfer function, different $\boldsymbol{k}$-modes corresponding to the same $k = |\boldsymbol{k}|$ bin have very different sensitivities. In order to get the best result, we need to take this into account when we calculate the spherically averaged power spectrum. As before, we use inverse noise-variance weighting to achieve this, giving us the following estimate for the unbiased spherically averaged pseudo-cross-spectrum:

$$\tilde{C}_{k_i} \equiv \frac{1}{\sum_{|\boldsymbol{k}|\in k_i} w_{\boldsymbol{k}}} \sum_{|\boldsymbol{k}|\in k_i} w_{\boldsymbol{k}} \tilde{C}_{\boldsymbol{k}}, \tag{23}$$

where $w_{\boldsymbol{k}} \equiv 1\big/\sigma^2_{\tilde{C}_{\boldsymbol{k}}}$ and k_i denotes the ith k-bin.

For simplicity, we only calculate the spherical average of the cross-spectra that have already been cylindrically averaged and binned. This means that we use the bin centers of the $k_\parallel \times k_\perp$ bins to represent all the modes in the bin, which means that a few $\boldsymbol{k}$-modes get shifted back or forth by one bin in the spherically averaged cross-spectrum. Since there are no sharp

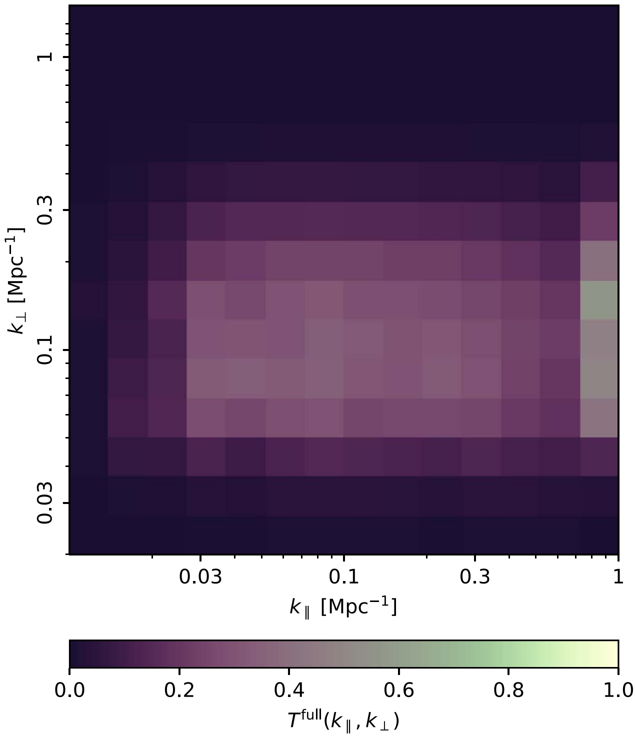


Figure 5. Full transfer function.

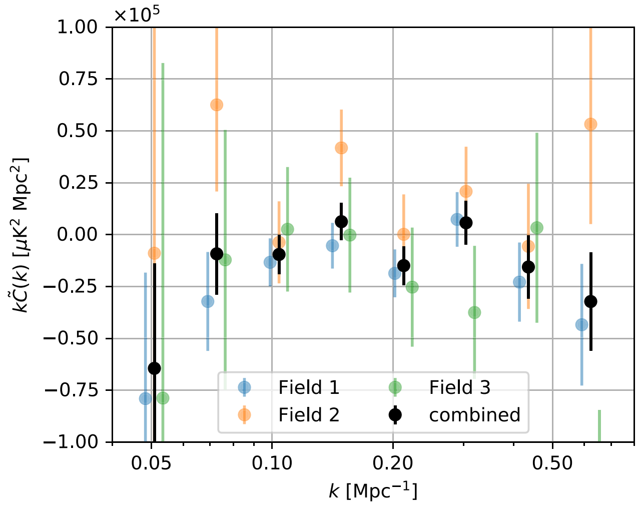


Figure 6. Spherically averaged mean pseudo-cross-spectra for CES observations of Field 1 (blue), Field 2 (orange), and Field 3 (green). These spectra were generated from all the accepted data using the FPXS cross-spectrum statistic. In addition, the full transfer function has been applied, to debias the signal estimate. The data points from the different fields are offset slightly in k from their actual values to make them easier to distinguish.

features in the signal power spectrum, this bias is modest, and not very important for the first-season analysis.

3. Power Spectrum Results

As described in Foss et al. (2022), after the COMAP time-ordered data have been filtered and calibrated, and the bad observations have been removed, the cleaned data set is compressed into a set of 3D maps. We make separate maps for the Lissajous scans and the CES scans, since these tend to have different systematics and statistical properties.

3.1. FPXS Results

We estimate separate cross-spectra for the Lissajous and CES data, for each of the three CO fields that we have observed (Foss et al. 2022). Since we found clear excess power in the Lissajous spectra, we do not include them in the main results, and we will here focus on the CES data. The power spectrum results for the Lissajous data are presented in Appendix A.

We split the data in two parts, according to the elevation of the observations, and use the FPXS method on these two sets of feed maps in order to minimize systematics. We also calculate a χ^2 statistic for each of the 16×15 different feed–feed cross-spectra,[16] $\tilde{C}^{ij}_{\boldsymbol{k}_i}$.

Based on these χ^2 statistics, denoted $\chi^2_{C(\boldsymbol{k})}$, we decided to reject all the spectra involving feed 8, since they showed very clear excesses in almost all spectra. This reduced the amount of data by 12.5% for all fields. We also saw clear structure in several of the spectra involving the low-elevation data from feeds 16 and 17 in the Field 1 results. This led us to remove all spectra involving these feeds from the low-elevation data set for Field 1, thereby increasing the data loss to 24.2% for this field. In addition to the spectra that were removed by hand, we also rejected all spectra with more than a 5σ excess in $\chi^2_{C(\boldsymbol{k})}$ before we calculated the FPXS mean spectrum.

In the automatic 5σ cut, the fractions of the remaining spectra that were removed for the CES data were given by 1/182 for Field 1, 159/210 for Field 2, and 65/210 for Field 3. As discussed in Foss et al. (2022), the fact that such large fractions of data were removed at this stage (especially for Fields 2 and 3) suggests that large improvements in sensitivity can be achieved in the future, if we can identify the data affected by systematic errors at an earlier stage of the pipeline.

The resulting spherically averaged pseudo-cross-spectra are shown in Figure 6. We see that the results for the CES data appear to be largely flat, with fluctuations that are consistent with our white noise estimate. This demonstrates that we are in fact averaging down the noise, as expected for uncorrelated noise, and that the various potential systematic errors are suppressed to a level below the noise. At this point in the COMAP survey, this is a key outcome, given that our fiducial theoretical model predicts a signal on the order of $kP_{\rm CO} \sim 10^3\ \mu{\rm K}^2\ {\rm Mpc}^2$ at our target redshift (Chung et al. 2022), well below the noise level shown here. This signal estimate is highly uncertain, however, and as discussed in Chung et al. (2022) these data already rule out some of the most optimistic models.

Combining these data points into a single measurement of the average CO power spectrum over the range $k = 0.051$–0.62 Mpc^{-1}, we get

$$P_{\rm CO}(k) = -2.7 \pm 1.7 \times 10^4 \mu{\rm K}^2{\rm Mpc}^3. \quad (24)$$

This estimate is based on the pseudospectrum, and, as discussed in Section 2.2, it is a somewhat biased estimate of the signal, but should be a conservative estimate if used as an upper limit, as we do in Chung et al. (2022). This is the first direct 3D constraint on the clustering component of the CO(1–0) power spectrum in the literature.

[16] As discussed in Foss et al. (2022), all data from feeds 4, 6, and 7 are rejected at an earlier stage of data selection. This leaves the data from 16 of the full 19 feeds.

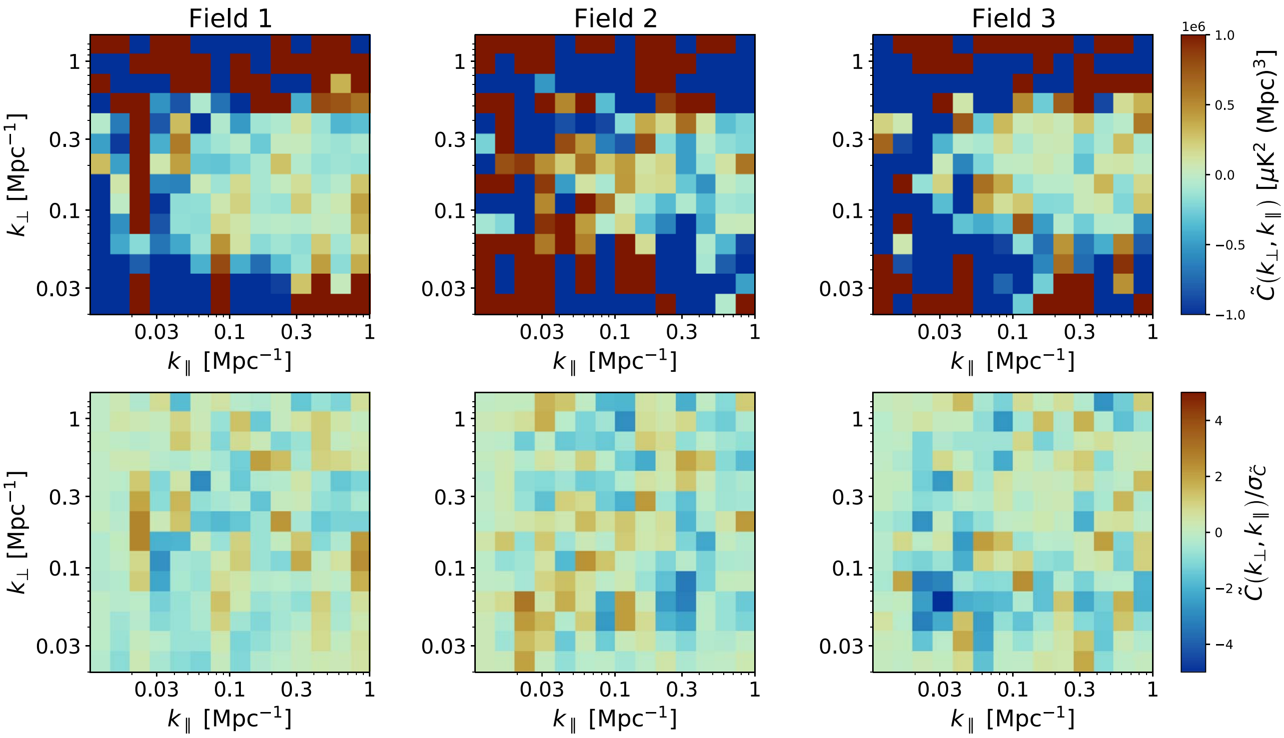


Figure 7. Cylindrically averaged mean pseudo-cross-spectra for CES observations (top row). The second row shows the spectra divided by the corresponding white noise uncertainties.

Figure 7 shows the corresponding cylindrically averaged power spectrum in $k_\parallel \times k_\perp$ space. We see that the noise blows up very quickly as we move away from the region where the transfer function peaks. This illustrates the importance of taking the transfer function into account during the spherical averaging if you want maximum sensitivity in the 1D spectrum. In the region where we have appreciable sensitivity, the spectra look consistent with white noise, as they do for the 1D spectra. The bottom row shows the spectra divided by their corresponding white noise uncertainties, to better illustrate what happens at all scales.

3.2. Null Tests

Given that the current data appear to be largely consistent with white noise, the importance of null tests is less critical than it would be if a potential detection had been made. Still, null tests are a useful consistency check, and they may be useful in identifying and highlighting specific systematic errors, as well as potentially provide hints regarding the nature of the Lissajous excesses.

In order to get a sensitive set of null tests, we can calculate cross-spectra between maps of our different fields. In these null tests, any systematic related to standing waves or from residual large-scale ground contamination could still show up, while the signal should not contribute at all. Moreover, as long as we center the fields appropriately, they are roughly as sensitive as our original spectra, while other null tests are typically less sensitive. We therefore perform the same kind of FPXS power spectrum estimation as for the main results, but use the high-elevation data from one field and the low-elevation data from another field. In this way, we obtain two null tests per field pair, one where the first field uses the low-elevation data while the second field uses the high-elevation data (denoted A), and another pair (denoted B) where the first field uses the high-elevation data and the other the low-elevation data. This gives a total of 6 null tests for each scanning method.

Figure 8 shows the results from these calculations, and we see that the null spectra are consistent with white noise expectations for all the CES data. For the Lissajous data, however, we see that most of the null tests show large excesses in power, consistent with our interpretation of the main Lissajous data containing systematics.

Table 1 shows χ^2 statistics calculated from each of the null tests in Figure 8, as well as the single-field results from Figures 6 and 10, by combining the eight data points of each spectrum. Specifically, we calculate the "probability to exceed," which is defined as 1 minus the cumulative distribution function of the χ^2 distribution with the given number of degrees of freedom (here, 8). If the data are given by white noise, these statistics should be evenly distributed between 0% and 100%, while if we have excess power present, then the values should tend to be small. The results in the table support and quantify the statements that we made above, that the CES data looks consistent with white noise and the Lissajous data has clear power excesses present.

Although it is hard to interpret these values precisely, the fact that we see a clear excess in most of the Lissajous spectra made by combining maps from different fields suggests that whatever the systematic effect that gives rise to this excess, it needs to be common to all fields.

4. Conclusion

In this paper, we have introduced FPXS as a robust method for estimating the CO signal power spectrum from 3D intensity

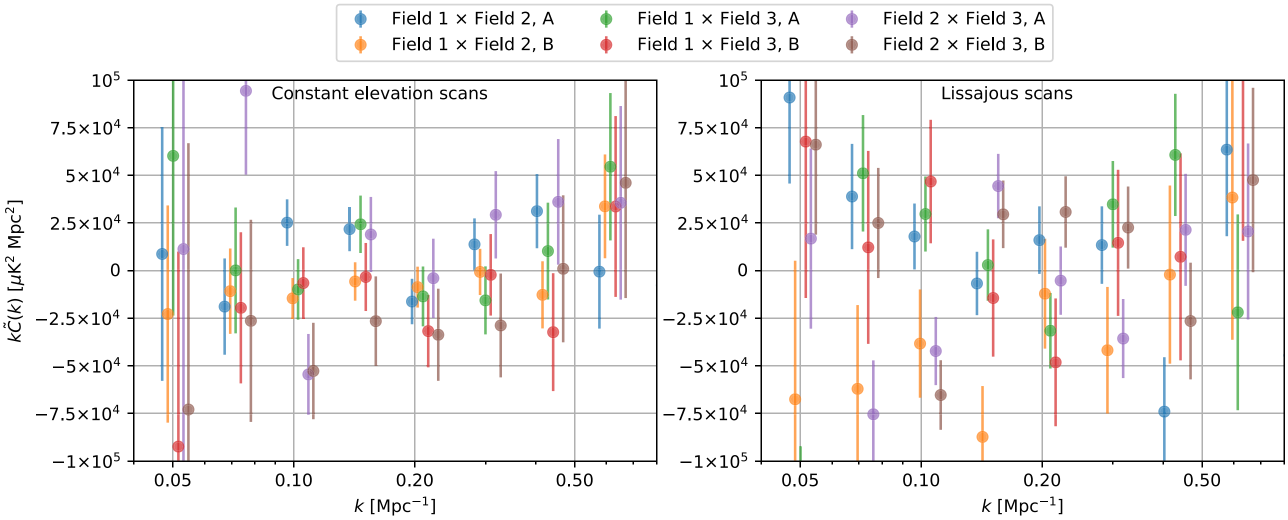


Figure 8. FPXS spectra of maps from different fields. Here, A denotes cross-spectra of the low-elevation map from the first field and the high-elevation map from the second field, while B denotes the opposite combination. The data points from the different spectra are offset slightly in k from their actual values to make them easier to distinguish.

Table 1
χ^2 Statistics from Science Results (Left) and Null Tests Using All the Different Field Combinations (Right)

	χ^2, Probability to Exceed									
Fields	All	Field 1	Field 2	Field 3	1 × 2, A	1 × 2, B	1 × 3, A	1 × 3, B	2 × 3, A	2 × 3, B
CES	33%	17%	30%	52%	9%	73%	52%	69%	5%	28%
Lissajous	0.02%	0.1%	3%	72%	3%	3%	0.5%	58%	0.3%	0.3%

Note. These values were calculated by combining the data points shown in Figures 6, 8, and 10 into a single χ^2 value for each spectrum.

maps produced by the COMAP data analysis pipeline. We have discussed how to account for signal loss due to both filtering and beam smoothing, and we have estimated their magnitudes for the first-year COMAP observations with simulations. Computing the FPXS from the actual COMAP data, we find that the current data set is consistent with white noise for constant-elevation scan data, and the uncertainties average down with time, as expected for ideal data. Equivalently, these results suggest that all systematic errors are lower than the white noise level in our main sensitivity range.

In contrast, the FPXS results from the Lissajous scan data show clear signs of systematic errors. Further modeling and analysis work is required before these data can be used for astrophysical analysis.

Null tests largely seem to be consistent with our main results, with all the CES null tests being consistent with white noise, while most of the Lissajous null tests show clear excesses, supporting our assumption that the excesses seen in the main Lissajous spectra are the result of systematics.

Future analysis will involve explicit estimation of the mode-mixing matrix (see Appendix D), to undo the mode-mixing effect and present unbiased power spectrum estimates. We can also increase our sensitivity (by up to a factor of $\sqrt{2}$) by splitting the data into more than the current low- and high-elevation sets.

This material is based upon work supported by the National Science Foundation, under grant Nos. 1517108, 1517288, 1517598, 1518282, and 1910999, and by the Keck Institute for Space Studies, under "The First Billion Years: A Technical Development Program for Spectral Line Observations." Parts of the work were carried out at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under a contract with the National Aeronautics and Space Administration, and funded through the internal Research and Technology Development program.

D.T.C. is supported by a CITA/Dunlap Institute postdoctoral fellowship. The Dunlap Institute is funded through an endowment established by the David Dunlap family and the University of Toronto. C.D. acknowledges support from an STFC Consolidated Grant (ST/P000649/1).

J.B., H.K.E., M.K.F., H.T.I., J.G.S.L., M.R., N.O.S., D.W., and I.K.W. acknowledge support from the Research Council of Norway through grants 251328 and 274990, and from the European Research Council (ERC) under the Horizon 2020 Research and Innovation Program (grant agreement No. 819478, COSMOGLOBE). J.G. acknowledges support from the University of Miami and is grateful to Hugh Medrano for assistance with cryostat design. S.H. acknowledges support from an STFC Consolidated Grant (ST/P000649/1). J.K. is supported by a Robert A. Millikan Fellowship from Caltech. At JPL, we are grateful to Mary Soria for assembly work on the amplifier modules and to Jose Velasco, Ezra Long, and Jim Bowen for the use of their amplifier test facilities. H.P. acknowledges support from the Swiss National Science Foundation through Ambizione Grant PZ00P2_179934. P.C.B. is supported by the James Arthur Postdoctoral Fellowship. We thank Isu Ravi for her contributions to the warm electronics and antenna–driven characterization. The scientific color maps `roma` and `tokyo` (Crameri 2021) are used in this study to prevent visual distortion of the data and exclusion of readers with color-vision deficiencies (Crameri et al. 2020).

Finally, we want to thank the anonymous referee, whose comments and suggestions have helped to improve and clarify this manuscript.

Software: Matplotlib (Hunter 2007); Astropy, a community-developed core Python package for astronomy (Astropy Collaboration et al. 2013).

Appendix A
Lissajous Power Spectrum Results

When looking at the Lissajous data, we found clear excess power in the spectra. For this reason, we do not use any Lissajous data for our final science results, but include the Lissajous power spectra here for completeness. We use the exact same approach for the Lissajous data as we did for the CES data. We derive a separate pipeline transfer function for the Lissajous data (see Figure 9), since the properties of these scans are a bit different from the CES scans. As noted in Foss et al. (2022), we find that the transfer function for the Lissajous data preserves a bit more large-scale angular structure than the CES one, but it does not make much qualitative difference in terms of the complete transfer function. In the automatic 5σ cut based on the $\chi^2_{C(\boldsymbol{k})}$ statistics, the fractions of remaining spectra that were removed for the Lissajous data were given by $132/182$ for Field 1, $92/210$ for Field 2, and $109/210$ for Field 3.

The Lissajous FPXS results for the spherically averaged power spectrum are shown in Figure 10. In contrast to the CES data, these data do not appear equally well behaved. Here, we see clear signs of excess power on large scales in both Field 1 and Field 2. These excesses suggest that large-scale systematic errors are still present for the Lissajous scans, and may be caused by either residual atmospheric variations or ground pickup from the far sidelobes, for instance.

These residuals are even more prominent when considering the 2D $k_\parallel \times k_\perp$ power spectrum, as shown in Figure 11. Here, we see some clear regions exhibiting systematic power excess. This is seen most clearly in the second row of the figure, which shows the power spectrum divided by the expected white noise fluctuations, thus corresponding to power measured in units of standard deviation. In particular, for Field 3, we see a bright region on the largest angular scales, and on scales between $k_\parallel \sim 0.03$–0.1 Mpc^{-1} in the frequency direction. We also see fairly bright regions at around $k_\perp = 0.2$ Mpc^{-1} and between $k_\parallel \sim 0.06$–0.6 Mpc^{-1}in the Field 1 data, which is right in the middle of our most sensitive region.

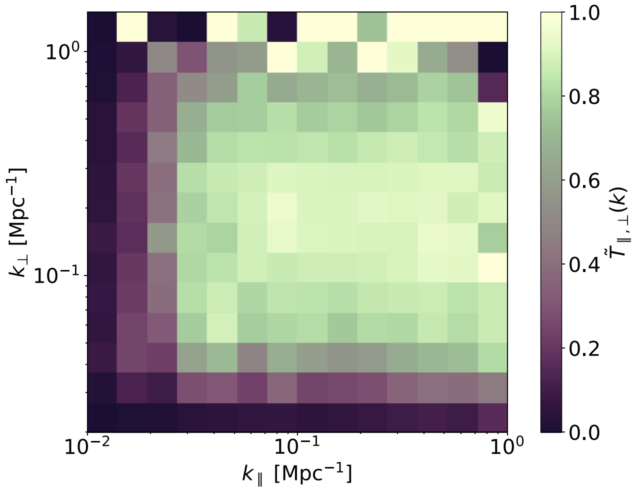


Figure 9. Pipeline transfer function for the cylindrically averaged power spectrum for the Lissajous scans. This transfer function is based on a single signal realization and roughly 3 hr of data.

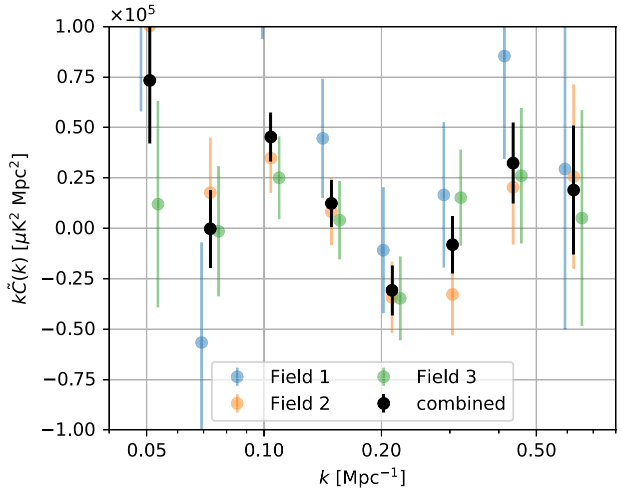


Figure 10. Spherically averaged mean pseudo-cross-spectra for the Lissajous observations of Field 1 (blue), Field 2 (orange), and Field 3 (green). These spectra were generated from all the accepted data using the FPXS cross-spectrum statistic. In addition, the full transfer function has been applied, to debias the signal estimate. The data points from the different fields are offset slightly in k from their actual values to make them easier to distinguish.

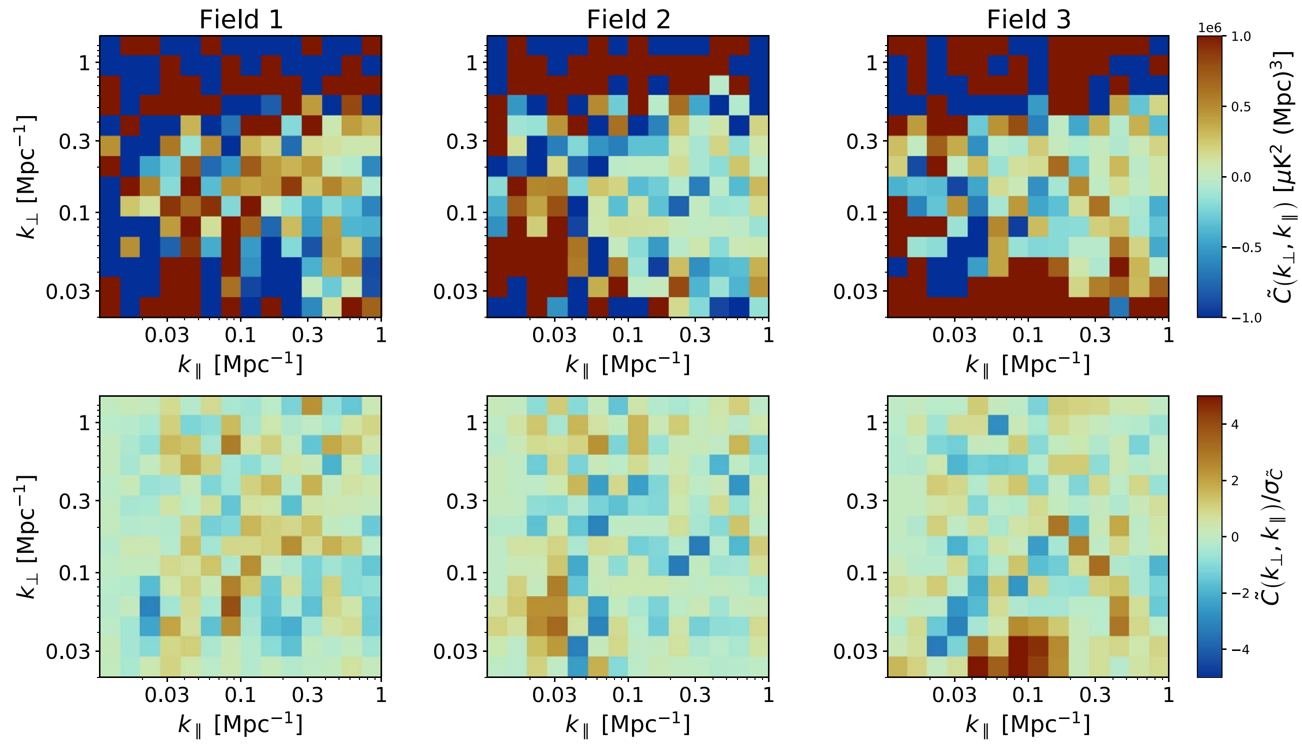


Figure 11. Cylindrically averaged mean pseudo-cross-spectra for the Lissajous observations (top row). The second row shows the spectra divided by the corresponding white noise uncertainties.

Appendix B Fourier Conventions

In this Appendix, we present the conventions for the discrete Fourier transformations used in this paper. All the conventions are consistent with the default conventions in NumPy's (Harris et al. 2020) FFT library. The forward transformation is given by

$$f_l = \sum_{m=0}^{n-1} x_m \exp\left(-2\pi i \frac{ml}{n}\right), \quad l = 0, \cdots, n-1,$$

where x_m are the discrete values of the function in real space and f_l are the Fourier coefficients. The inverse transformation is then given by

$$x_m = \frac{1}{n}\sum_{l=0}^{n-1} f_l \exp\left(2\pi i \frac{ml}{n}\right).$$

We define the physical wavenumber as follows:

$$\begin{aligned} k &\equiv \frac{2\pi j}{\Delta x n}, \quad j \in \left\{-\frac{n}{2}, \cdots, -1, 0, 1, \cdots, \frac{n}{2}\right\} \\ &= 2\pi \cdot \mathrm{np.fft.fftfreq}(n, \Delta x). \end{aligned}$$

Appendix C Definition of Cosmological Map Grid

Since Fourier transforms require a rectangular grid, we assume that the 3D temperature maps can be approximated by a rectangular grid in comoving cosmological parameters. We assume that all the voxels have the same shape and size as the middle voxel at redshift $z_{\mathrm{mid}} \approx 2.9$.

The comoving length corresponding to an angular separation $\delta\theta$, for a given redshift z, is given by

$$\delta l_\perp = r(z)\delta\theta = \delta\theta \int_0^z \frac{cdz'}{H(z')}, \tag{C1}$$

where $r(z)$ is the comoving distance traveled by light emitted from redshift z to us.

The comoving radial distance corresponding to a small redshift interval $\delta z = z_1 - z_2 = \nu_0/\nu_1^{\mathrm{obs}} - \nu_0/\nu_2^{\mathrm{obs}} \approx (1+z)^2 \delta \nu^{\mathrm{obs}}/\nu_0$, where $z_1 > z_2$, is given by

$$\delta l_\parallel = \int_{z_2}^{z_1} \frac{cdz}{H(z)} \approx \frac{c\delta z}{H(z)} \approx \frac{c}{H(z)} \frac{(1+z)^2 \delta\nu^{\mathrm{obs}}}{\nu_0}, \tag{C2}$$

where $\nu_0 \approx 115.27$ is the emission frequency of the CO $1 \rightarrow 0$ line we are studying and $\delta\nu^{\mathrm{obs}} = 31.25$ MHz is the resolution of our frequency bins.

Given a pixel width of $2'$, we then get the following voxel dimensions:

$$\delta l_\perp = 3.63 \text{ Mpc}, \tag{C3}$$

$$\delta l_\parallel = 4.26 \text{ Mpc}. \tag{C4}$$

Appendix D Mode Mixing and the Master Algorithm

In order to understand the mode-mixing effect, let us consider in more detail the Fourier transform of a weighted map:[17]

[17] We work in 2D here to save some indices; the generalization to 3D is straightforward.

$$\tilde{f}_{k_1k_2} = \sum_{m_1=0}^{n-1}\sum_{m_2=0}^{n-1} x_{m_1m_2} W_{m_1m_2} \times \exp\left(-2\pi i \frac{m_1k_1 + m_2k_2}{n}\right). \quad \text{(D1)}$$

Here, $x_{m_1m_2}$ is the map, $W_{m_1m_2}$ is the weight map, and $\tilde{f}_{k_1k_2}$ is the Fourier transform of the weighted map. We can insert the expression for the inverse Fourier transform of x and W,

$$\tilde{f}_{k_1k_2} = \frac{1}{n^4}\sum_{k_1'=0}^{n-1}\sum_{k_2'=0}^{n-1} f_{k_1'k_2'} \sum_{k_1''=0}^{n-1}\sum_{k_2''=0}^{n-1} f^W_{k_1''k_2''} \times \sum_{m_1=0}^{n-1}\sum_{m_2=0}^{n-1} \times\exp\left(-2\pi i\frac{m_1(k_1' + k_1'' - k_1) + m_2(k_2' + k_2'' - k_2)}{n}\right), \quad \text{(D2)}$$

where f and f^W are the Fourier transforms of x and W, respectively. Working through the algebra, we get

$$\tilde{f}_{k_1k_2} = \frac{1}{n^2}\sum_{k_1'=0}^{n-1}\sum_{k_2'=0}^{n-1} f_{k_1'k_2'} \sum_{k_1''=0}^{n-1}\sum_{k_2''=0}^{n-1} f^W_{k_1''k_2''}\, \delta_{k_1''(k_1-k_1')\%n, k_2''(k_2-k_2')\%n}$$

$$\tilde{f}_{k_1k_2} = \frac{1}{n^2}\sum_{k_1'=0}^{n-1}\sum_{k_2'=0}^{n-1} f_{k_1'k_2'} f^W_{(k_1-k_1')\%n(k_2-k_2')\%n}$$

$$\tilde{f}_{k_1k_2} = \sum_{k_1'=0}^{n-1}\sum_{k_2'=0}^{n-1} f_{k_1'k_2'} \underbrace{\frac{1}{n^2} f^W_{(k_1-k_1')\%n(k_2-k_2')\%n}}_{\equiv K_{k_1,k_2,k_1',k_2'}}, \quad \text{(D3)}$$

where % denotes the modulo operation and we have defined the mode-mixing amplitude $K_{\boldsymbol{k},\boldsymbol{k}'}$.

Adopting vector notation, we may now write the pseudospectrum as follows:

$$\tilde{P}(\boldsymbol{k}) = \frac{V_{\text{vox}}}{N_{\text{vox}}}\langle \tilde{f}_{\boldsymbol{k}} \tilde{f}^*_{\boldsymbol{k}}\rangle \quad \text{(D4)}$$

$$= \frac{V_{\text{vox}}}{N_{\text{vox}}}\frac{1}{n^{2D}}\sum_{\boldsymbol{k}'}\sum_{\boldsymbol{k}''}\langle f_{\boldsymbol{k}'} f^*_{\boldsymbol{k}''}\rangle K_{\boldsymbol{k},\boldsymbol{k}'} K^*_{\boldsymbol{k},\boldsymbol{k}''} \quad \text{(D5)}$$

$$= \frac{1}{n^{2D}}\sum_{\boldsymbol{k}'}\sum_{\boldsymbol{k}''} P(\boldsymbol{k}')\delta_{\boldsymbol{k}',\boldsymbol{k}''} K_{\boldsymbol{k},\boldsymbol{k}'} K^*_{\boldsymbol{k},\boldsymbol{k}''} \quad \text{(D6)}$$

$$= \sum_{\boldsymbol{k}'} P(\boldsymbol{k}') \underbrace{\frac{1}{n^{2D}}|K_{\boldsymbol{k},\boldsymbol{k}'}|^2}_{M_{\boldsymbol{k},\boldsymbol{k}'}}, \quad \text{(D7)}$$

where D is the number of dimensions of the map and we have defined the mode-mixing matrix $M_{\boldsymbol{k},\boldsymbol{k}'}$. We see that the autospectrum and the pseudospectrum are related by a linear transformation, so all the information in one is also there in the other.

Within the cosmic microwave background field, accounting for mode mixing by explicitly calculating and inverting $M_{\boldsymbol{k},\boldsymbol{k}'}$ is often referred to as the MASTER algorithm (Hivon et al. 2002; Leung et al. 2022). Doing this requires that we calculate the mode mixing between each Fourier mode and all the other Fourier modes, so for 3D maps this scales poorly with the map dimensions. On the other hand, the algorithm parallelizes trivially, and the matrix must only be computed once for a given weight map, after which the same operation may be applied efficiently to any number of simulations. Whether this is feasible depends on the details of the individual use case. Some methods exist in the literature to approximate this procedure in a faster way; see, e.g., Louis et al. (2020).

ORCID iDs

Håvard T. Ihle https://orcid.org/0000-0003-3420-7766
Kieran A. Cleary https://orcid.org/0000-0002-8214-8265
Hans Kristian Eriksen https://orcid.org/0000-0003-2332-5281
Marie K. Foss https://orcid.org/0000-0001-8896-3159
Stuart E. Harper https://orcid.org/0000-0001-7911-5553
Junhan Kim https://orcid.org/0000-0002-4274-9373
Liju Philip https://orcid.org/0000-0001-7612-2379
Nils-Ole Stutzer https://orcid.org/0000-0001-5301-1377
Bade D. Uzgil https://orcid.org/0000-0001-8526-3464
Duncan J. Watts https://orcid.org/0000-0002-5437-6121
Ingunn Kathrine Wehus https://orcid.org/0000-0003-3821-7275
J. Richard Bond https://orcid.org/0000-0003-2358-9949
Patrick C. Breysse https://orcid.org/0000-0001-8382-5275
Dongwoo T. Chung https://orcid.org/0000-0003-2618-6504
Clive Dickinson https://orcid.org/0000-0002-0045-442X
Delaney A. Dunne https://orcid.org/0000-0002-5223-8315
Andrew I. Harris https://orcid.org/0000-0001-6159-9174
James W. Lamb https://orcid.org/0000-0002-5959-1285
Anthony C. S. Readhead https://orcid.org/0000-0001-9152-961X
Hamsa Padmanabhan https://orcid.org/0000-0002-8800-5740
Timothy J. Pearson https://orcid.org/0000-0001-5213-6231
Thomas J. Rennie https://orcid.org/0000-0002-1667-3897

References

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Battye, R. A., Davies, R. D., & Weller, J. 2004, MNRAS, 355, 1339
Bernal, J. L., Breysse, P. C., Gil-Marin, H., & Kovetz, E. D. 2019, PhRvD, 100, 123522
Breysse, P. C., Anderson, C. J., & Berger, P. 2019, PhRvL, 123, 231105
Breysse, P. C., Chung, D. T., Cleary, K. A., et al. 2022, ApJ, 933, 188
Breysse, P. C., Kovetz, E. D., Behroozi, P. S., Dai, L., & Kamionkowski, M. 2017, MNRAS, 467, 2996
Breysse, P. C., Kovetz, E. D., & Kamionkowski, M. 2014, MNRAS, 443, 3506
Carilli, C. L., & Walter, F. 2013, ARA&A, 51, 105
Chung, D. T. 2019, ApJ, 881, 149
Chung, D. T., Breysse, P. C., Cleary, K. A., et al. 2022, ApJ, 933, 186
Chung, D. T., Breysse, P. C., Ihle, H. T., et al. 2021, ApJ, 923, 188
Cleary, K. A., Borowska, J., Breysse, P. C., et al. 2022, ApJ, 933, 182
Crameri, F. 2021, Scientific colour maps, v7.0.0, Zenodo, doi:10.5281/zenodo.1243862
Crameri, F., Shephard, G. E., & Heron, P. J. 2020, NatCo, 11, 5444
Foss, M. K., Ihle, H. T., Borowska, J., et al. 2022, ApJ, 933, 184
Gong, Y., Chen, X., & Cooray, A. 2020, ApJ, 894, 152
Hamilton, A. J. S. 1998, in The Evolving Universe, ed. D. Hamilton (Dordrecht: Kluwer), 185
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, ApJ, 567, 2
Hunter, J. D. 2007, CSE, 9, 90
Ihle, H. T., Chung, D., Stein, G., et al. 2019, ApJ, 871, 75
Keating, G. K., Marrone, D. P., Bower, G. C., & Keenan, R. P. 2020, ApJ, 901, 141

Keenan, R. P., Keating, G. K., & Marrone, D. P. 2022, ApJ, 927, 161
Kovetz, E., Breysse, P. C., Lidz, A., et al. 2019, BAAS, 51, 101
Kovetz, E. D., Viero, M. P., Lidz, A., et al. 2017, arXiv:1709.09066
Lamb, J. W., Cleary, K. A., Woody, D. P., et al. 2022, ApJ, 933, 183
Leung, J. S. Y., Hartley, J., Nagy, J. M., et al. 2022, ApJ, 928, 109
Li, T. Y., Wechsler, R. H., Devaraj, K., & Church, S. E. 2016, ApJ, 817, 169
Lidz, A., Furlanetto, S. R., Oh, S. P., et al. 2011, ApJ, 741, 70
Loeb, A., & Wyithe, J. S. B. 2008, PhRvL, 100, 161301
Louis, T., Naess, S., Garrido, X., & Challinor, A. 2020, PhRvD, 102, 123538
Madau, P., Meiksin, A., & Rees, M. J. 1997, ApJ, 475, 429
Mertens, F. G., Mevius, M., Koopmans, L. V. E., et al. 2020, MNRAS, 493, 1662
Moradinezhad Dizgah, A., Nikakhtar, F., Keating, G. K., & Castorina, E. 2022, JCAP, 2022, 026
Peterson, J. B., Bandura, K., & Pen, U. L. 2006, arXiv:astro-ph/0606104
Pullen, A. R., Chang, T.-C., Doré, O., & Lidz, A. 2013, ApJ, 768, 15
Rennie, T. J., Harper, S. E., Dickinson, C., et al. 2022, ApJ, 933, 187
Sato-Polito, G., & Bernal, J. L. 2022, arXiv:2202.02330
Uzgil, B. D., Carilli, C., Lidz, A., et al. 2019, ApJ, 887, 37
Yang, S., Somerville, R. S., Pullen, A. R., et al. 2021, ApJ, 911, 132

Paper III

COMAP Pathfinder – Season 2 results I. Improved data selection and processing

Astronomy & Astrophysics manuscript no. aanda
June 18, 2024

COMAP Pathfinder – Season 2 results
I. Improved data selection and processing

J. G. S. Lunde[1,⋆], N.-O. Stutzer[1], P. C. Breysse[6,7], D. T. Chung[3,4,5], K. A. Cleary[2], D. A. Dunne[2], H. K. Eriksen[1], S. E. Harper[10], H. T. Ihle[1], J. W. Lamb[9], T. J. Pearson[2], L. Philip[12,16], I. K. Wehus[1], D. P. Woody[9], J. R. Bond[3,14,15], S. E. Church[17], T. Gaier[12], J. O. Gundersen[18], A. I. Harris[19], R. Hobbs[9], J. Kim[13], C. R. Lawrence[12], N. Murray[3,14,15], H. Padmanabhan[8], A. C. S. Readhead[2], T. J. Rennie[10,11], and D. Tolgay[3,14] (COMAP Collaboration)

1 Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, N-0315 Oslo, Norway
2 California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
3 Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada
4 Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
5 Department of Astronomy, Cornell University, Ithaca, NY 14853, USA
6 Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY, 10003, U.S.A
7 Department of Physics, Southern Methodist University, Dallas, TX 75275, USA
8 Departement de Physique Théorique, Universite de Genève, 24 Quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
9 Owens Valley Radio Observatory, California Institute of Technology, Big Pine, CA 93513, USA
10 Jodrell Bank Centre for Astrophysics, Alan Turing Building, Department of Physics and Astronomy, School of Natural Sciences, The University of Manchester, Oxford Road, Manchester, M13 9PL, U.K.
11 Department of Physics and Astronomy, University of British Columbia, Vancouver, BC Canada V6T 1Z1, Canada
12 Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109
13 Department of Physics, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
14 Department of Physics, University of Toronto, 60 St. George Street, Toronto, ON, M5S 1A7, Canada
15 David A. Dunlap Department of Astronomy, University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada
16 Brookhaven National Laboratory, Upton, NY 11973-5000
17 Kavli Institute for Particle Astrophysics and Cosmology and Physics Department, Stanford University, Stanford, CA 94305, USA
18 Department of Physics, University of Miami, 1320 Campo Sano Avenue, Coral Gables, FL 33146, USA
19 Department of Astronomy, University of Maryland, College Park, MD 20742

Received MM DD, YYYY; accepted MM DD, YYYY

ABSTRACT

The CO Mapping Array Project (COMAP) Pathfinder is performing line intensity mapping of CO emission to trace the distribution of unresolved galaxies at redshift $z \sim 3$. We present an improved version of the COMAP data processing pipeline and apply this to the first two seasons of observations. This analysis improves on the COMAP Early Science (ES) results in several key aspects. On the observational side, all second-season scans were made in constant-elevation mode, after noting that the previous Lissajous scans were associated with increased systematic errors; those scans accounted for 50% of the total Season 1 data volume. Secondly, all new observations were restricted to an elevation range of 35–65 degrees, to minimize sidelobe ground pickup. On the data processing side, more effective data cleaning in both the time- and map-domain has allowed us to eliminate all data-driven power spectrum-based cuts. This increases the overall data retention and reduces the risk of signal subtraction bias. On the other hand, due to the increased sensitivity, two new pointing-correlated systematic errors have emerged, and we introduce a new map-domain PCA filter to suppress these. Subtracting only 5 out of 256 PCA modes, we find that the standard deviation of the cleaned maps decreases by 67% on large angular scales, and after applying this filter, the maps appear consistent with instrumental noise. Combining all these improvements, we find that each hour of raw Season 2 observations yields on average 3.2 times more cleaned data compared to ES analysis. Combining this with the increase in raw observational hours, the effective amount of data available for high-level analysis is a factor of 8 higher than in ES. The resulting maps have reached an uncertainty of 25–50 μK per voxel, providing by far the strongest constraints on cosmological CO line emission published to date.

Key words. galaxies: high-redshift – radio lines: galaxies – diffuse radiation – methods: data analysis

⋆ e-mail: j.g.s.lunde@astro.uio.no

1. Introduction

Line intensity mapping (LIM) is an emerging observational technique in which the integrated spectral line emission from many unresolved galaxies is mapped in 3D as a tracer of cosmological large-scale structure (e.g., Kovetz et al. 2017, 2019). It represents a promising and complementary cosmological probe to, say, galaxy surveys and cosmic microwave background (CMB) observations. In particular, LIM offers the potential to survey vast cosmological volumes at high redshift in a manner that is sensitive to emission from the entire galaxy population, not just the brightest objects, as is the case for high-redshift galaxy surveys (Bernal & Kovetz 2022). The most widely studied emission line for LIM purposes is the 21-cm line of neutral hydrogen (Loeb & Zaldarriaga 2004; Bandura et al. 2014; Santos et al. 2017), which is the most abundant element in the universe, but other high-frequency emission lines also appear promising (Korngut et al. 2018; Pullen et al. 2023; Akeson et al. 2019; Crites et al. 2014; CCAT-Prime Collaboration et al. 2022; Vieira et al. 2020; Karkare et al. 2022; Fasano et al. 2024), in particular, due to their different and complementary physical origin, as well as lower levels of astrophysical confusion, Galactic foregrounds and radio frequency interference.

The CO Mapping Array Project (COMAP) represents one example of such an alternative approach and uses CO as the tracer species (see, e.g. Lidz et al. 2011; Pullen et al. 2013; Breysse et al. 2014). The COMAP Pathfinder instrument consists of a 19-feed[1] focal plane array observing at 26–34 GHz (Lamb et al. 2022), deployed on a 10.4 m Cassegrain telescope. This frequency range corresponds to a redshift of $z \sim 2.4$–3.4 for the CO(1–0) line, a period during the Epoch of Galaxy Assembly (Li et al. 2016). The Pathfinder instrument started observing in 2019, and COMAP has previously published results from the first year of data, named Season 1, in our Early Science (ES) publications (Cleary et al. 2022; Lamb et al. 2022; Foss et al. 2022; Ihle et al. 2022; Chung et al. 2022; Rennie et al. 2022; Breysse et al. 2022; Dunne et al. 2024). These ES results provided the tightest constraints on the CO power spectrum in the clustering regime published to date. Since the release of the ES results, the COMAP Pathfinder instrument has continued to observe, while also implementing many important lessons learned from Season 1 both in terms of observing strategy and data processing methodology. Combining the observations from both Seasons 1 and 2, and improving the data analysis procedure, the new results improve upon the ES analysis by almost an order of magnitude in terms of power spectrum sensitivity.

This paper is the first of the COMAP Season 2 paper series, and here we present the low-level data analysis pipeline and map-level results derived from the full COMAP data set as of the end of Season 2 (November 2023). This work builds on the corresponding Season 1 effort as summarized by Ihle et al. (2022). The corresponding power spectrum and null-test results are presented by Stutzer et al. (2024), while Chung et al. (2024) discuss their cosmological implications in terms of structure formation constraints. In parallel with the Season 2 CO observations, the COMAP Pathfinder continues to survey the Galactic plane, with the latest results focusing on the Lambda Orionis region (Harper et al. 2024).

The remainder of this paper is structured as follows: In Sect. 2 we summarize the changes made to the observational strategy in Season 2 and provide an overview of the current status of data collection and accumulated volume. In Sect. 3 we summarize our time-ordered data (TOD) pipeline with a focus on the changes since ES. In Sect. 4 we study the statistical properties of the spectral maps produced by this pipeline while paying particular attention to our new map-domain principal component analysis (PCA) filtering and the systematic errors that this filter is designed to mitigate. In Sect. 5 we present the current data selection methodology and discuss the resulting improvements in terms of data retention in the time, map, and power spectrum domains. In Sect. 6 we present updated end-to-end pipeline transfer function estimates and discuss their generality with respect to non-linear filtering. Finally, we summarize and conclude in Sect. 7.

[1] We refer to a full detector chain as a "feed".

2. Data collection and observing strategy

Table 1 shows the raw on-sky integration time per season. Here we see that COMAP Season 1 included 5,200 on-sky observation hours collected from May 2019 to August 2020, while the second COMAP season included 12,300 hours collected between November 2020 and November 2023. In these publications, we present results based on a total on-sky integration time of 17,500 hours, a 3.4-fold increase in raw data volume compared to the ES publications.

We start by reviewing the changes made to the data collection and observing strategy before and during Season 2. Most of these changes came as direct responses to important lessons learned during the Season 1 data analysis and aimed to increase the net mapping speed, although one was necessitated due to mechanical telescope issues during Season 2. Overall, these changes were highly successful, and Season 2 has much higher data retention than Season 1, which will be discussed in Sect. 5. The most important changes in the Season 2 observing strategy are the following:

1. Observations were restricted to an elevation range of 35°–65° in order to reduce the impact of ground pick-up via the telescope's sidelobe response. As discussed by Ihle et al. (2022), the gradient of the ground pickup changes quickly at both lower and higher elevations, and the corresponding observations were therefore discarded in the Season 1 analysis; in Season 2 we avoid these problematic elevations altogether.
2. Similarly, Lissajous scans were abandoned in favor of solely using constant elevation scans (CES), since Foss et al. (2022) found elevation-dependent systematic errors associated with the former.
3. The two frequency detector sub-bands, which previously covered disjoint ranges of 26–30 GHz and 30–34 GHz (Lamb et al. 2022), were widened slightly, such that they now overlap; this mitigates data loss due to aliasing near the band edges.
4. The acceleration of the azimuth drive was halved to increase the longevity of the drive mechanism, which started to show evidence of mechanical wear.

The latter two changes were only implemented in the second half of Season 2, and mark the beginning of what we refer to as Season 2b. These changes are now discussed in greater detail.

2.1. Restricting the elevation range

Sidelobe pickup of the ground is a potentially worrisome systematic error for COMAP, especially since it is likely to be pointing-correlated. Even though ground pickup is primarily correlated with pointing in alt-azimuthal coordinates, the daily repeating pointing pattern of COMAP means there will still be a strong

Table 1. Overview of COMAP observation season definitions

Name	Dates	Observing hours	Notes
Season 1	05/2019 - 08/2020	5,200	Data published in Early Science. Contains 50% Lissajous, 50% CES.
Season 2a	11/2020 - 04/2022	7,900	100% CES from this point forward.
Season 2b	05/2022 - 11/2023	4,400	After azimuth drive slowdown and sampler frequency change.

Table 2. Season 2 is split into two sub-season, respectively denoted 2a and 2b, as the telescope scanning speed was significantly reduced in May 2022 for mechanical reasons.

correlation in equatorial coordinates. Analysis of Season 1 observations (Foss et al. 2022), which ranged from ∼30° to ∼75° elevation, showed pointing-correlated systematic errors at the highest and lowest elevations.

To study this effect in greater detail, we developed a set of antenna beam pattern simulations using GRASP[2] for the COMAP telescope (Lamb et al. 2022), and these showed the presence of four sidelobes resulting from the four secondary support legs (SSL), with each sidelobe offset by ∼65° from the pointing center. These simulations were convolved with the horizon elevation profile at the telescope site, and the results from these calculations are shown in Fig. 1. This figure clearly shows that Fields 2 and 3 experience a sharp change in ground pickup around 65°–70° elevation, as one SSL sidelobe transitions between ground and sky. At very low elevations the ground contribution also varies rapidly for all fields as two of the other SSL sidelobes approach the horizon. While the low-level TOD pipeline removes the absolute signal offset per scan, gradients in the sidelobe pickup over the duration of a scan still lead to signal contamination. We have therefore restricted our observations to the elevation range of 35°–65°, where one SSL sidelobe remains pointed at the ground, and the other three SSL sidelobes are safely pointing at the sky, leaving us with a nearly constant ground pickup. This change incurred little loss in observational efficiency, as almost all allocated observational time outside the new range could be reallocated to other fields within the range.

2.2. Abandoning Lissajous scans

The first season of observations contained an even distribution of Lissajous and constant elevation scans, with the aim of exploring the strengths and weaknesses of each. The main strength of the Lissajous scanning technique is that it provides excellent cross-linking by observing each pixel from many angles, which is useful for suppressing correlated noise with a destriper or maximum likelihood mapmaker. The main drawback of this observing mode is that the telescope elevation varies during a single scan, resulting in varying atmosphere and ground pick-up contributions. In contrast, the telescope elevation remains fixed during a constant elevation scan (CES), producing a simpler pick-up contribution although with somewhat worse cross-linking properties.

When analyzing the Season 1 power spectra resulting from each of the two observing modes, Ihle et al. (2022) found that the Lissajous data both produced a highly significant power spectrum, especially on larger scales and failed key null tests. The CES scans, on the other hand, produced a power spectrum consistent with zero, and passed the same null tests. We therefore concluded that the significance in the Lissajous power spectrum was due to residual systematic errors. Additionally, the main

[2] https://www.ticra.com/software/grasp/

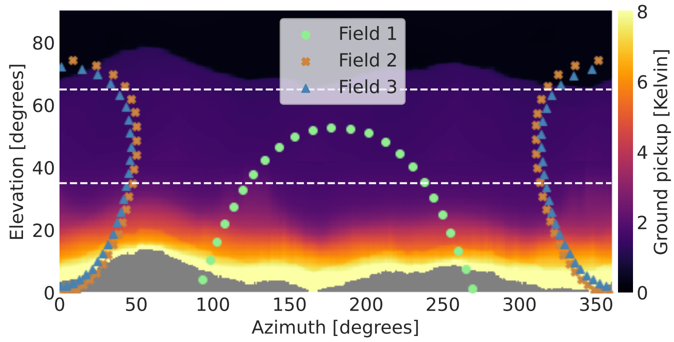


Fig. 1. Approximate sidelobe ground pickup as a function of az/el pointing, simulated by convolving a beam (simulated using GRASP) with the horizon profile (shown in gray) at the telescope site. The paths of the three fields across the sky, in half-hour intervals, are shown, as well as the Season 2 elevation limits at 35°–65°. These limits ensure minimal ground pickup gradient across the field paths.

advantage of Lissajous scanning, namely better cross-linking, proved virtually irrelevant because of a particular feature of the COMAP instrument: because all frequency channels in a single COMAP sideband are processed through the same backend, the instrumental $1/f$ gain fluctuations are extremely tightly correlated across each sideband. As a result, the low-level TOD pipeline is capable of simultaneously removing virtually all correlated noise from both gain and atmosphere by common-mode subtraction (see Sect. 3.5). At our current sensitivity levels, we therefore find no need to employ a complex mapmaking algorithm that fully exploits cross-linking observations, but we can rather use a computationally faster binned mapmaker (Foss et al. 2022). After Season 1 we therefore concluded that there was no strong motivation to continue with Lissajous scans, and in Season 2 we employ solely CES.

2.3. Widening of frequency bands for aliasing mitigation

As discussed in detail by Lamb et al. (2022), the COMAP instrument exhibits a small but non-negligible level of signal aliasing near the edge of each sideband. In the Season 1 analysis, this was accounted for simply by excluding the channels with aliasing power from other channels suppressed by less than 15 dB. In total, 8 % of the total COMAP bandwidth is lost due to this effect, and this leads to gaps in the middle of the COMAP frequency range. Both the origin of the problem and its ultimate solution were known before the Season 1 observations started (Lamb et al. 2022), but this took time to implement.

Band-pass filters applied after the first downconversion and low-pass filters applied after the second downconversion allow significant power above the Nyquist frequency into the sampler. This is then aliased into the 0–2.0 GHz observing baseband,

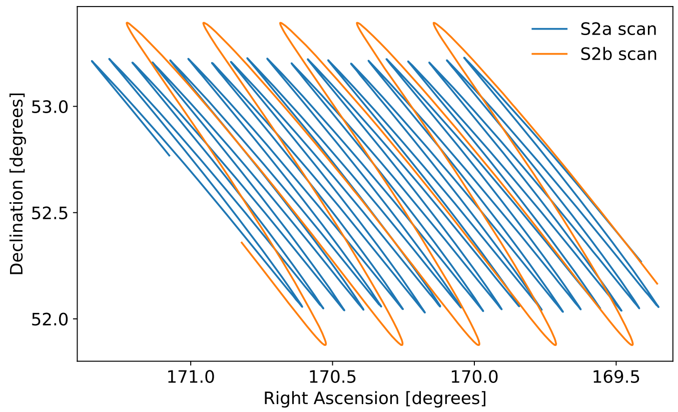


Fig. 2. Comparison of the faster pointing pattern from a Season 2a scan, and the slower pointing pattern of a Season 2b scan. Both patterns show a 5.5-minute constant elevation scan, as the field drifts across.

requiring the contaminated channels to be excised. By increasing the sampling frequency from 4.000 GHz to 4.250 GHz, the Nyquist frequency is raised to 2.125 GHz, closer to the filter edges. Not only is the amount of aliased power reduced, it is also folded into frequencies above the nominal width of each 2 GHz observing band. The existing samplers were able to accommodate the higher clock speed, but the field-programmable gate array (FPGA) code had to be carefully tuned to reliably process the data. This was finally implemented from the start of Season 2b, and the aliased power is now shifted outside the nominal range of each band, such that the affected channels can be discarded without any loss in frequency coverage. The number of channels across the total frequency range is still 4096, so the "native" Season 2b channels are 2.075 MHz wide, up from 1.953 MHz.

2.4. Azimuth drive slowdown

It became clear during Season 2 that the performance of the telescope's azimuth drives had degraded, owing to wear and tear on the drive mechanisms caused by the telescope's high accelerations. In order to protect the drives from damage, the analog acceleration limit was reduced until the stress was judged by its audible signature to be acceptable. Though not carefully quantified, this was about an order of magnitude change, and the minimum time for a scan is therefore about a factor of three less. Additionally, the maximum velocity was reduced by a factor of two in the drive software.

Figure 2 illustrates the old (Season 2a) and new (Season 2b) pointing patterns, with the new pattern being slightly wider and around four times slower. The new realized pointing pattern is now also closer to sinusoidal since the drives are better able to 'keep up' with the sinusoidal pattern of the commanded position, due to the slower velocity.

With the new actually sinusoidal scanning pattern, the integration time is less uniform across each field in each observation, as the telescope spends more time pointing near the edges of the field than it previously did. However, co-adding across the observing season does smooth out the uneven integration time, based on the receiver field of view (each of the 19 feeds observes the sky at a position that is offset from the others) and field rotation (the telescope observes the fields at different angles as they move across the sky).

2.5. Data storage

With 19 feeds, 4096 native frequency channels, and a sampling rate of about 50 samples/sec, COMAP collects 56 GB/hour of raw 24-bit integer data, stored losslessly as 32-bit floats. The full set of these raw data (combined with telescope housekeeping), named "Level 1", currently spans about 800 TB of disk space. These data are then filtered by our TOD pipeline into so-called "Level 2" data (Foss et al. 2022), in which a key step is to co-add the native frequency channels to 31.125 MHz width. These downsampled channels form the basis of the higher-level map-making and power spectrum algorithms. The total amount of Level 2 data is about 50 TB. Both Level 1 and Level 2 files are now losslessly compressed using the GZIP algorithm (Gailly & Adler 1992), which achieves average compression factors of 2.4 and 1.4, respectively, reducing the effective sizes of the two datasets to 330 TB and 35 TB. The lower compression factor of the filtered data is expected because the filtering leaves the data much closer to white noise, and therefore with a much higher entropy.

The files are stored in the HDF5 format (Koranne 2011), which allows seamless integration of compression. Both compression and decompression happen automatically when writing to and reading from each file. GZIP is also a relatively fast compression algorithm, taking roughly one hour to compress each hour of COMAP data on a typical single CPU core. Decompression takes a few minutes per hour of data, which is an insignificant proportion of the total pipeline runtime. HDF5 also allows for arbitrary chunking of datasets before compression. Chunking aids in optimizing performance since the Level 1 files consist of entire observations (1 hour), but the current TOD pipeline implementation (see Sect. 3.9) reads only individual scans of 5 minutes each. We partition the data into chunks of 1-minute intervals to minimize redundant decompression when accessing single scans. Other compression algorithms have been tested, and some, such as lzma[3], achieve up to a 20 % higher compression factor on our data. They are, however, also much slower at both compression and decompression, and they interface less easily with HDF5.

3. The COMAP TOD pipeline

This section lays out the COMAP time-domain pipeline, named `l2gen`, focusing on the changes from the first generation pipeline, which is described in detail by Foss et al. (2022). The pipeline has been entirely re-implemented (see Sect. 3.9) for performance and maintainability reasons but remains mathematically similar. Figure 3 shows a flowchart of the entire COMAP pipeline, of which `l2gen` is the first step.

The purpose of the TOD pipeline is to convert raw detector readout (Level 1 files) into calibrated time-domain data (Level 2 files) while performing two key operations: substantially reducing correlated noise and systematic errors, and calibrating to brightness temperature units. COMAP uses a filter-and-bin pipeline, meaning that we perform as much data cleaning as possible in the time domain, before binning the data into maps with naïve noise-weighting. This leaves us with a biased pseudo-power spectrum, which can be corrected by estimating the pipeline transfer function (Sect. 6).

The following sections will explain the main filters in the TOD pipeline. The normalization (Sec. 3.3), $1/f$ filter (Sec. 3.5), calibration and downsampling (Sect. 3.8) steps remain unchanged from the ES pipeline, and are briefly summarized for

[3] https://tukaani.org/xz/

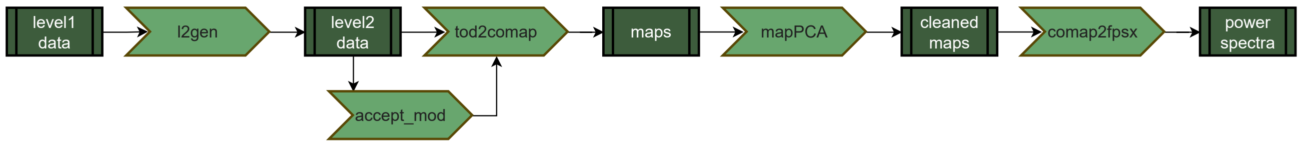


Fig. 3. Flowchart of the COMAP pipeline, from raw Level1 data to final power spectra. Data products are shown as darker boxes and pipeline code as lighter arrows. `l2gen` performs the time-domain filtering, turning raw data into cleaned Level 2 files. `accept_mod` performs scan-level data selection on cleaned data. `tod2comap` is a simple binned mapmaker. `mapPCA` performs a map-level PCA filtering. Finally, `comap2fpsx` calculates power spectra as described by Stutzer et al. (2024).

completeness. We will denote the data at different stages of the pipeline as $d_{\nu,t}^{\mathrm{name}}$, where the ν, t subscript indicates data with both frequency and time dependence.

3.1. System temperature calculation

The first step of the pipeline is to calculate the system temperature T_ν^{sys} of each channel in the TOD. At the beginning and end of each observation, a vane of microwave-absorbing material of known temperature is temporarily moved into the field of view of the entire feed array. The measured power from this "hot load", P_ν^{hot}, and the temperature of the vane, T^{hot}, are interpolated between calibrations to the center of each scan. Power from a "cold load", P_ν^{cold}, is calculated as the average power of individual sky scans. The system temperature is then calculated as (see Foss et al. (2022) for details)

$$T_\nu^{\mathrm{sys}} = \frac{T^{\mathrm{hot}} - T^{\mathrm{CMB}}}{P_\nu^{\mathrm{hot}}/P_\nu^{\mathrm{cold}} - 1}. \tag{1}$$

under the approximation that the ground, sky, and telescope share the same temperature.

3.2. Pre-pipeline masking

In ES, `l2gen` performed all frequency channel masking towards the end of the pipeline. While some masks are data-driven (specifically driven by the filtered data), others are not. We now apply the latter category of masks prior to the filters, to improve the filtering effectiveness. These are

- masking of channels that have consistently been found to be correlated with systematic errors, and have been manually flagged to always be masked;
- for data gathered before May 2022, masking of channels with significant aliased power, as outlined in Sect. 2.3;
- masking of channels with system temperature spikes, as outlined in the section below.

3.2.1. System temperature spikes

The system noise temperature, T_ν^{sys}, for each feed's receiver chain has a series of spikes at specific frequencies, believed to result from an interaction between the polarizers and the corrugated feed horn (Lamb et al. 2022). The spikes are known to be associated with certain systematic errors, and the affected frequency channels are therefore masked out. The ES analysis used a static system temperature threshold of 85 K for masking, but the new version of `l2gen` applies a 400 channel-wide running median kernel to the data and masks all frequencies with a noise temperature of more than 5 K above the median. We repeat the running median fit and threshold operation once on the masked data, to reduce the impact of the spikes on the fit. The second iteration uses a kernel width of 150 channels. The final running median and threshold are illustrated in Fig. 4. As the system temperature can vary quite a lot across the 4 GHz range, this method fits the spikes themselves more tightly, while avoiding cutting away regions of elevated but spike-free system temperature.

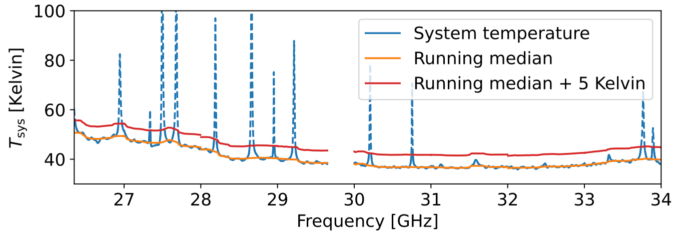


Fig. 4. Example of T^{sys} spike masking by running median. The system temperature is shown in blue, the running median in orange, and the 5 Kelvin cut above the running median in red.

The spike frequencies vary from feed to feed, and we are therefore not left with gaps in the redshift coverage of the final 3D maps. On average, we mask 6% of all frequency channels this way. However, because the affected channels are, by definition, more noisy, this only results in a loss of 3% of the sensitivity.

3.3. Normalization

The first filtering step in the TOD pipeline is to normalize the Level 1 data by dividing by a low-pass filtered version of the data and subtracting 1. The filter can be written as

$$d_{\nu,t}^{\mathrm{norm}} = \frac{d_{\nu,t}^{\mathrm{raw}}}{\langle d^{\mathrm{raw}} \rangle_\nu} - 1, \quad \langle d^{\mathrm{raw}} \rangle_\nu = \mathcal{F}^{-1}\{W \cdot \mathcal{F}\{d_{\nu,t}^{\mathrm{raw}}\}\}, \qquad W = \left(1 + \left(\frac{f}{f_{\mathrm{knee}}}\right)^\alpha\right)^{-1}, \tag{2}$$

where $\mathcal{F}$ is the Fourier transform and $\mathcal{F}^{-1}$ is the inverse Fourier transform, both performed on the time-dimension of the data, and W is a low pass filter in the Fourier domain, with a spectral slope $\alpha = -4$, and a knee frequency $f_{\mathrm{knee}} = 0.01$ Hz. This has the effect of removing all modes on 100 s timescales and longer. As the COMAP telescope crosses the entire field in 5–20 seconds, the normalization has minimal impact on the sky signal in the scanning direction, but heavily suppresses the signal perpendicular to the scanning direction, as the fields take 5–7 minutes to drift across.

The filter is performed per frequency channel, and the primary purpose of the normalization is to remove the channel-to-channel gain variations, making channels more comparable. After applying this filter, the white noise level in each channel will now be the same and common-mode $1/f$ gain fluctuations will be flat across frequency. As a secondary consideration, the normalization also removes slow-running atmospheric and gain

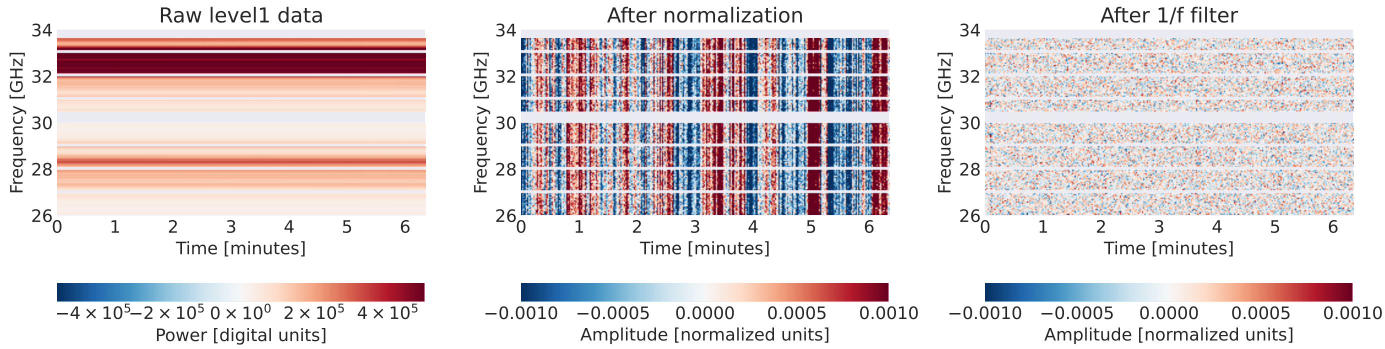


Fig. 5. Illustration of the effect of TOD normalization and $1/f$ filtering for a single feed and scan. The raw Level 1 data (left) are dominated by frequency-dependent gain variations, which correspond to the instrumental bandpass. After normalization (middle), the signal is dominated by common-mode gain fluctuations. Finally, after the $1/f$ filter is applied (right), the common-mode $1/f$ contribution has been suppressed, and the data are dominated by white noise. The horizontal gray stripes indicate channels that were masked by the pipeline. All three stages happen before absolute calibration, and the amplitudes are therefore given in arbitrary units.

fluctuations on timescales greater than 100 s. The effect of the filter can be seen in the first two panels of Fig. 5, which shows the TOD of a single scan in 2D before and after the normalization. Before the normalization, the frequency-dependent gain dominates, and the time variations are invisible. After normalization, the data in each channel fluctuates around zero.

3.4. Azimuth filter

Next, we fit and subtract a linear function in azimuth, to reduce the impact of pointing-correlated systematic errors, first and foremost being ground pickup by the telescope sidelobes. This filter can be written as

$$d_{\nu,t}^{\mathrm{az}} = d_{\nu,t}^{\mathrm{norm}} - a_\nu \cdot \mathrm{az}_t \tag{3}$$

where a_ν is fitted to the data per frequency, and az_t is the azimuth pointing of the telescope. Unlike in the ES pipeline, this filter is now fitted independently for when the telescope is moving eastward and westward, to mitigate some directional systematic effects we have seen.

In Season 1, we also employed Lissajous scans, meaning that an elevation term was also present in this equation. As we now only observe in constant elevation mode, this term falls away.

3.5. $1/f$ gain fluctuation filter

After normalization, the data are dominated primarily by gain, and secondarily by atmospheric fluctuations, and both are strongly correlated on longer timescales. Although the normalization suppresses power on all timescales longer than 100 seconds, we observe that common-mode noise still dominates the total noise budget down to ∼1 s timescales.

To suppress this correlated noise, we apply a specific $1/f$ filter[4] by exploiting the simple frequency behavior of the gain and atmosphere fluctuations. After we have normalized the data, the amplitude of the gain fluctuations is the same across all frequency channels, although fluctuating in time. The contribution of the atmosphere, an approximate continuum source in temperature, becomes almost linear in frequency in normalized units. We therefore fit and subtract a first-order polynomial across frequency for every time step;

$$d_{\nu,t}^{1/\mathrm{f}} = d_{\nu,t}^{\mathrm{point}} - (c_t^0 + c_t^1 \cdot \nu), \tag{4}$$

where c_t^0 and c_t^1 are coefficients fitted to the data each time step. To ensure that the atmosphere can be well approximated as a continuum, we fit a separate linear polynomial to the 1024 channels of each of the four 2 GHz sidebands.

This simple technique is remarkably efficient at removing $1/f$ noise, and we observe that the correlated noise is suppressed by several orders of magnitude, after which white noise dominates the uncertainty budget. This is illustrated in the last two panels of Fig. 5. After the normalization, the signal is completely dominated by common-mode $1/f$ noise. After the $1/f$ filter, the correlated noise is effectively suppressed, and we are left with almost pure white noise, as can be seen in the right panel, and further discussed in Sect. 3.10.

3.6. PCA filtering

Principal Component Analysis (PCA) is a common and powerful technique for dimensionality reduction (Pearson 1901). Given a data matrix $m_{\nu,t}$, PCA produces an ordered basis w_t^k for the columns of $m_{\nu,t}$, called the *principal components* of $m_{\nu,t}$. The component *amplitude* can then be calculated by re-projecting the components into the matrix, as $a_\nu^k = m_{\nu,t} \cdot w_t^k$. For our purposes, $m_{\nu,t}$ is the TOD, with frequencies as rows, and time-samples as columns. The ordering of the principal components w_t^k is such that the earlier components capture as much of the variance in the columns of $m_{\nu,t}$ as possible, and for any selected number of components N_{comp}, the following expression is minimized:

$$\sum_{\nu,t} \left(m_{\nu,t} - \sum_{k=1}^{N_{\mathrm{comp}}} a_\nu^k w_t^k \right)^2. \tag{5}$$

In other words, PCA provides a compressed version of $m_{\nu,t}$, which approximates $m_{\nu,t}$ as the sum of an ordered set of outer products[5]

$$m_{\nu,t} \approx \sum_{k=1}^{N_{\mathrm{comp}}} a_\nu^k w_t^k. \tag{6}$$

[4] The filter is referred to as the polynomial filter in our ES publications.

[5] PCA has several equivalent interpretations and ways of solving for the principal components. The principal components are, among other things, the eigenvectors of the covariance matrix of $m_{\nu,t}$. This is how we introduced the PCA in our ES publications. It is, however, both a slow way of solving for the PCA components in practice and not the best interpretation for our purposes.

PCA is often employed on a dataset where the rows are interpreted as different observations, and the columns are the multi-dimensional features of these data. However, this is not a natural interpretation for our purposes, and it makes more sense to simply look at PCA as a way of compressing a 2D matrix as a sum of outer products – we have no special distinction between columns and rows, and could equivalently have solved for the PCA of the transpose of $m_{\nu,t}$, which would swap a_ν and w_t.

A PCA is often performed because one is interested in keeping the leading components, as these contain much of the information in the data. However, we *subtract* the leading components, because many systematic errors naturally decompose well into an outer product of a frequency vector and a time vector, while the CO signal does not (and is very weak in a single scan).

In practice, we solve for the leading principal components using a singular value decomposition algorithm (Halko et al. 2011), and then calculate the amplitudes as stated above. The N_{comp} leading components are then subtracted from the TODs, leaving us with the filtered data

$$d^{\mathrm{PCA}}_{\nu,t} = m_{\nu,t} - \sum_{k=1}^{N_{\mathrm{comp}}} a^k_\nu w^k_t. \tag{7}$$

The Season 2 COMAP pipeline employs two time-domain PCA filters, one of which was present in ES. In the following subsection, we introduce both filters and then explain how to decide the number of leading components, N_{comp}, to subtract from the data.

The process of calculating the principal components and subtracting them from the data constitutes a non-linear operation on the data. This has the advantage of being much more versatile against systematic errors that are difficult to model using linear filters, but the disadvantage is a more complicated impact on the CO signal itself. This will be further discussed in Sect. 6.4, where our analysis shows that the PCA filter behaves linearly with respect to any sufficiently weak signal, and that, at the scan-level, all plausible CO models (Chung et al. 2021) are sufficiently weak by several orders of magnitude.

3.6.1. The all-feed PCA filter

The *all-feed PCA filter*, which was also present in the ES pipeline, collapses the 19 feeds onto the frequency axis of the matrix, producing a data matrix $m_{\nu,t}$ of the $1/f$-filtered data with a shape of $(N_{\mathrm{feed}}\ N_{\mathrm{freq}}, N_{\mathrm{TOD}}) = (19 \times 4096, \sim 20\,000)$ for a scan with N_{feed} feeds, N_{freq} frequency channels, and N_{TOD} time samples. The PCA algorithm outlined above is then performed on this matrix. Combining the feed and frequency dimensions means that a feature in the data will primarily only be picked up by the filter if it is common (in the TOD) across all 19 feeds. This is primarily the case for any atmospheric contributions, and potentially standing waves that originate from the optics common to all feeds. It is, however, certainly not the case for the CO signal, which will be virtually unaffected by this filter.

Figure 6 shows a typical strong component w_t picked up by this filter in the top panel, with the component amplitudes a_ν for three selected feeds in the bottom panel. This systematic error appears to be consistent with a standing wave, i.e., appearing as a ripple whose amplitude varies with frequency.

3.6.2. The per-feed PCA filter

The new *per-feed PCA filter* has been implemented to combat systematic errors that vary from feed to feed. This filter employs the PCA algorithm outlined above on each individual feed and is performed on the output of the all-feed PCA filter. Additionally, we found that downsampling the data matrix (using inverse variance noise weighing) by a factor of 16 in the frequency direction before performing the PCA increased its ability to pick up structures in the data. The resulting data matrix $m_{\nu,t}$ gets the shape $(N_{\mathrm{freq}}/16, N_{\mathrm{TOD}}) = (256, \sim 20\,000)$ for each feed. The downsampling is only used when solving for the time-domain components w_t, and the full data matrix is used when calculating the frequency amplitudes, a_ν.

Targeting each feed individually makes us more susceptible to CO signal loss, but the low signal-to-noise ratio (SNR), combined with the fact that the CO signal can not be naturally decomposed into an outer product, makes the impact on the CO signal itself minimal. This filter appears to primarily remove components consistent with standing waves from the individual optics and electronics of each feed.

Figure 7 demonstrates a large-scale fluctuation picked up by the per-feed PCA filter for feed 14. The origin of the shape of the systematic error is unknown, but it constitutes a typical behavior of the component and amplitude of this filter.

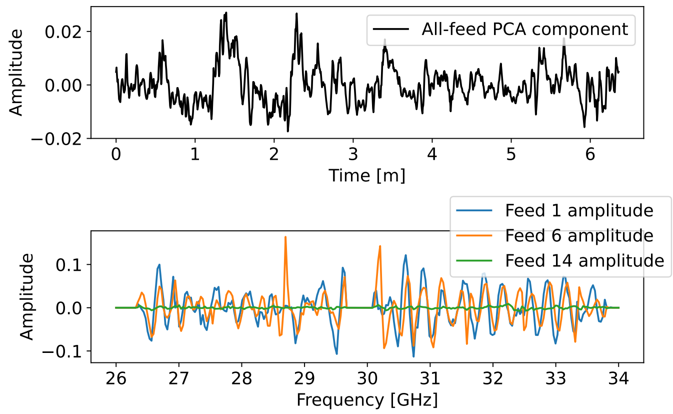


Fig. 6. (*Top:*) Illustration of a significant all-feed PCA component. (*Bottom:*) Frequency amplitudes of the above component across three randomly selected feeds. The amplitudes show clear signs of a feed-common standing wave. Since the filter is applied to normalized data, the y-axis amplitudes are unitless.

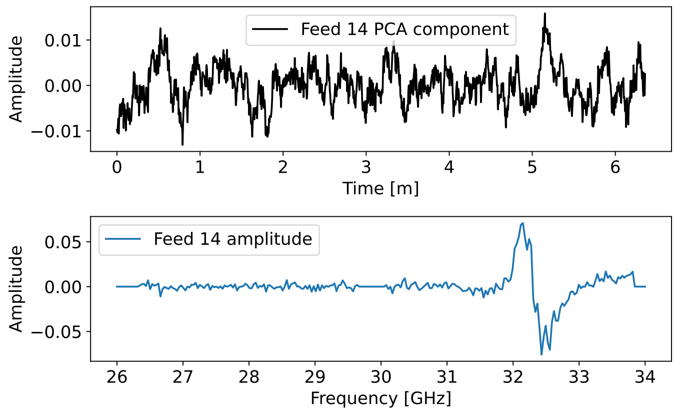


Fig. 7. Same as Fig. 6, but for a component of the per-feed PCA filter, where the time-domain PCA component belongs to feed 14 alone. This component has picked up some frequency structure localized around 32 GHz.

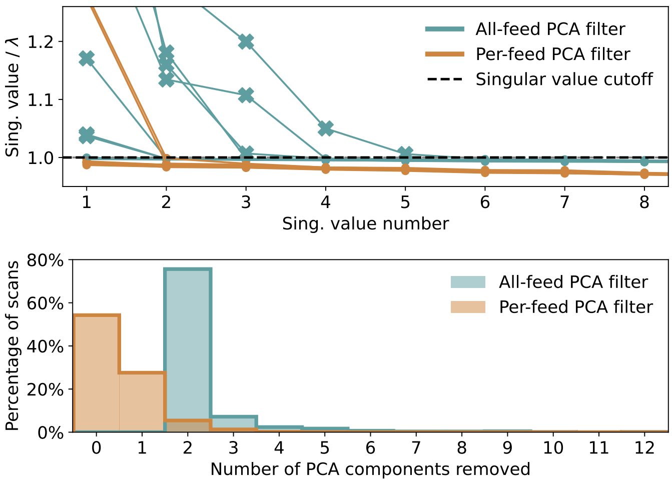


Fig. 8. *(Top:)* Largest singular values of the all-feed and per-feed PCA filters, divided by λ, for a random selection of ten scans. PCA components with relative values above one are removed from the data and are marked with crosses in this plot. *(Bottom:)* Number of PCA components subtracted across all scans. At least two components are always subtracted by the all-feed PCA filter.

3.6.3. Number of components

In the ES pipeline, the number of PCA components was fixed at four for the all-feed filter, and the per-feed filter did not exist. We now dynamically determine the required number of components for each filter, per scan. This allows us to use more components when needed, removing more systematic errors, and fewer when not needed, incurring a smaller loss of CO signal.

We subtract principal components until the components are indistinguishable from white noise, which can be inferred from the singular values of each component. Let λ be the expectation value of the largest singular value of a (N, P) Gaussian noise matrix (see Appendix A for how this value is derived). We subtract principal components until we reach a singular value below λ. However, for safe measure, we always subtract a minimum of 2 components for the all-feed PCA filter, as the signal impact of this filter is minimal.

Figure 8 shows typical singular values, relative to λ, for a random selection of scans, and a histogram of the number of components employed across all scans. The average number of PCA components subtracted is 2.3 and 0.5 for the all-feed and per-feed PCA, respectively; the most common number of components subtracted is the minimum allowed in each case: two and zero. The top part of the figure also demonstrates that there is a sharp transition between the singular values of the components which actually pick up meaningful features from the signal, and the remaining noise components.

3.7. Data-inferred frequency masking

After the PCA filters, we perform dynamic masking of frequency channels identified from the filtered TOD. This mainly consists of masking groups of channels that have substantially higher correlations between each other than expected from white noise, explained in more detail in Foss et al. (2022). This is assumed to be caused by substantial residuals of gain fluctuations or atmospheric signal. We do this by calculating the correlation matrix between the frequency channels, and then looking at boxes and stripes of various sizes across this matrix, performing χ^2 tests on their white noise consistency. We also mask individual channels with a standard deviation significantly higher than expected from the radiometer equation.

After the frequency masking, the $1/f$ filter and PCA filters are reapplied to the data, to ensure that their performance was not degraded by misbehaving channels. The normalization and pointing filters do not need to be reapplied, as they work independently on each frequency channel.

3.8. Calibration and downsampling

The final step of the pipeline is to calibrate the data to temperature units and decrease the frequency resolution. After the normalization step, the data are in arbitrary normalized units. Using the system temperature calculated in Sect. 3.1 we calibrate each channel of the data,

$$d_{\nu,t}^{\mathrm{cal}} = T_\nu^{\mathrm{sys}} d_{\nu,t}^{\mathrm{PCA}}. \tag{8}$$

Finally, we downsample the frequency channels from 4096 native channels to 256 science channels. The Seasons 1 and 2a frequency channels are 1.953 MHz wide, while they are 2.075 MHz wide in Season 2b, after the change in sampling frequency (Sect. 2.3). In both cases, the channels are downsampled to a grid of 31.25 MHz, which exactly corresponds to a factor 16 downsampling for the older data. For the newer data, either 15 or 16 native channels will contribute to each science channel, decided by their center frequency. The downsampling is performed with inverse-variance weighting, using the system temperature as uncertainty.

3.9. Implementation and performance

While the ES pipeline was written in Fortran, the Season 2 pipeline has been rewritten from scratch to run in Python. Performance-critical sections are either written in C++ and executed using the Ctypes package or employ optimized Python packages like SciPy. Overall, the serial performance is similar to the ES pipeline, but the Season 2 pipeline employs a more fine-grained and optimal MPI+OpenMP parallelization, making it much faster on systems without a very large memory-to-core ratio.

The pipeline is run on a small local cluster of 16 E7-8870v3 CPUs, with a total of 288 cores, in about a week of wall time, totaling around 40,000 CPU hours for the full COMAP dataset. The time-domain processing dominates this runtime, with a typical scan taking around 20–25 minutes to process on a single CPU core.

3.10. Time-domain results

The Level 2 TODs outputted by `l2gen` are assumed to be almost completely uncorrelated in both time and frequency dimensions, such that the TOD are well approximated as white noise. To quantify the correlations in the time domain we calculate the temporal power spectrum of each individual channel for all scans. Figure 9 shows this power spectrum averaged over both scans and frequencies, compared both to the equivalent power spectrum of un-filtered Level 1 data, and to that of a TOD obtained by injecting pure white noise in place of our real data into the TOD pipeline.

Since the pipeline filtering removes more data on longer timescales, the white noise simulation (red) gradually deviates from a flat power spectrum on longer timescales, until it falls

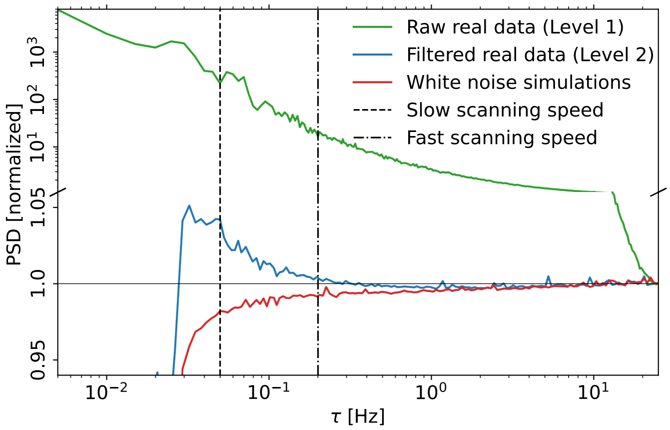


Fig. 9. Average temporal power spectrum of unfiltered (green) and filtered scans (blue), compared to correspondingly filtered white noise simulations (red). The y-axis is broken at 1.06 and is logarithmic above this. The data are averaged over scans, feeds, and frequencies, and are normalized with respect to the highest k-bin. For context, the old and new scanning frequencies (once across the field) are shown as vertical lines (dot-dashed and dashed, respectively).

rapidly at timescales below ~ 0.03 Hz due to the high-pass normalization performed in the pipeline. The power spectrum for the filtered real data (blue) also follows the same trend on short timescales, but then increases on timescales around $k = 0.2$ Hz, due to small residual $1/f$ gain fluctuations and atmosphere remaining after the processing. Below ~ 0.03 Hz, this power spectrum again falls rapidly due to the high-pass filter. The power spectrum obtained from raw data which have not gone through any filtering (green), simply increases on longer timescales as expected, due to $1/f$ gain and atmospheric fluctuations.

The difference between the blue and red spectra shows the residual correlated noise left in the data. While there is some residual correlated noise, it is an insignificant fraction of our final noise budget. We find that our real filtered data have a standard deviation only 1.7 % higher than that of the filtered white noise. Compared to the amount of power we see in the unfiltered data, we see that our pipeline is very efficient at suppressing correlated noise.

4. Mapmaking & map domain filtering

4.1. The COMAP mapmaker

COMAP employs a simple binned inverse-variance noise-weighted mapmaker, identical to the one in ES (Foss et al. 2022). This can be written as

$$m_{\nu,p} = \frac{\sum_{t\in p} d_{\nu,p}/\sigma^2_{\nu,t}}{\sum_{t\in p} 1/\sigma^2_{\nu,t}}, \tag{9}$$

where $m_{\nu,p}$ is an individual map voxel[6], $d_{\nu,t}$ represents the time-domain data over time samples, $\sigma^2_{\nu,t}$ is the time-domain white noise uncertainty, and $t \in p$ means all time-samples t which observes pixel p. We assume that the white noise uncertainty $\sigma_{\nu,t}$ is constant (for a single feed and frequency) over the duration of a scan, and calculate it per scan as

$$\sigma_\nu = \sqrt{\frac{\mathrm{Var}(d_{\nu,t} - d_{\nu,t-1})}{2}}, \tag{10}$$

then let $\sigma_{\nu,t} = \sigma_\nu$ for all time-samples within the scan. This value is also binned into maps, and is used as the uncertainty estimate of the maps throughout the rest of the analysis:

$$\sigma^2_{\nu,p} = \frac{1}{\sum_{t\in p} 1/\sigma^2_t}. \tag{11}$$

In practice, we calculate per-feed maps, both because the map-PCA filter (introduced in the next section) is performed on per-feed maps, and because the cross-spectrum algorithm utilizes groups of feed-maps.

The reason for not using more sophisticated mapmaking schemes, like destriping (Keihänen et al. 2010) or maximum likelihood mapmaking (Tegmark 1997), is partially of necessity – the COMAP TOD dataset is many hundreds of TB, making iterative algorithms difficult. However, the TOD pipeline has also proven remarkably capable of cleaning most unwanted systematic errors from the data, especially correlated $1/f$ noise, as we saw from Fig. 9. As the main purpose of more sophisticated mapmaking techniques is dealing with correlated noise, COMAP is served well with a simple binned mapmaker.

The mapmaking algorithm is identical to the one used for the ES analysis. However, as with `l2gen` the actual implementation has been rewritten from scratch in Python and C++, with a focus on optimal parallelization and utilization of both MPI and OpenMP.

4.2. Map-domain PCA filtering

The pipeline now employs a PCA filtering step also in the map domain, in addition to the one we apply at the TOD level. This technique is almost entirely analogous to the PCA foreground subtraction often employed in 21cm LIM experiments (Chang et al. 2010; Masui et al. 2013; Anderson et al. 2018), although we do not employ it to subtract foregrounds. The primary purpose of this filter is to mitigate a couple of pointing-correlated systematic errors (see the next subsection) which proved challenging to remove entirely in the time domain. The method is similar to the TOD PCA algorithm from Sect. 3.6, but instead of having the TOD data matrix $d_{\nu,t}$, we have a map $m_{\nu,p}$ with one frequency and one (flattened) pixel dimension. The data matrix then gets the shape $(N_\nu, N_p) = (256,\ 14400)$, with $N_\nu = 256$ being the number of frequency channels in the map and $N_p = 120 \cdot 120 = 14400$ the number of pixels in each frequency slice (although many of the pixels in each individual feed-map are never observed).

The technique we employ here is technically a slight generalization of the PCA problem, as we want to weigh individual voxels by their uncertainty when solving for the components and amplitudes[7]. This is not possible in the regular PCA framework without also morphing the modes one is trying to fit, as we explain in Appendix B. As shown in Sect. 3.6, the first principal component w^0_p and its amplitude a^0_ν are the vectors that minimize the value of the expression

$$\sum_{\nu,p} (m_{\nu,p} - a^0_\nu w^0_p)^2. \tag{12}$$

[6] A voxel here is the 3D equivalent of a pixel, with two angular dimensions and a redshift/frequency dimension. Here we separate the voxel dimensions into frequencies ν and pixels p.

[7] For the time-domain PCA, it was enough to weight individual channels, with all time-samples in that channel sharing the same weight. This can be done with a normal PCA, as we show in Appendix B.

In other words, they are the two vectors such that their outer product explains as much of the variance in $m_{\nu,p}$ as possible. This formulation of the PCA makes it obvious how to generalize the problem to include weighting for individual matrix elements: we can simply minimize the following sum,

$$\sum_{\nu,p} \frac{(m_{\nu,p} - a^0_\nu w^0_p)^2}{\sigma^2_{\nu,p}}, \tag{13}$$

where $\sigma_{\nu,p}$ is the uncertainty in each voxel. Minimizing Eq. (13) gives us the vectors a^0_ν and w^0_p for which the outer product $a^0_\nu w^0_p$ explains as much of the variance in $d_{\nu,p}$ as possible, weighted by $\sigma_{\nu,p}$. The resulting map $d_{\nu,p} - a^0_\nu w^0_p$ represents the filtered map. The process can then be repeated any number of times, solving for and subtracting a new set of vectors. We minimize the expression in Eq. (13) iteratively with an algorithm outlined in Appendix B, where we also explain why this is not equivalent to simply performing the PCA on a noise-weighted map $d_{\nu,p}/\sigma_{\nu,p}$. Due to the large similarity of our technique to a regular PCA, we simply refer to this filter as a PCA filter.

As for the TOD PCA filter, a selected number of components are subtracted from the data maps

$$m^{\mathrm{mPCA}}_{\nu,p} = m_{\nu,p} - \sum_{i=1}^{N_{\mathrm{comp}}} a^k_\nu v^k_p \tag{14}$$

This filtering is performed per feed, as the systematic errors outlined in the next subsection manifest differently in different feeds. We have chosen $N_{\mathrm{comp}} = 5$, which will be further explained in Sect. 4.3.3. Because the COMAP scanning strategy stayed the same throughout Seasons 1 and 2a but changed with the azimuth slowdown of Season 2b, we apply the map-PCA separately to the former and latter, as the pointing-correlated effects we are trying to remove might also be different.

4.3. Newly-discovered systematic effects

The two most prominent new systematic errors discovered in the second season of observations have been dubbed the "turn-around" and "start-of-scan" effects. They have in common that they are difficult to model in the time domain, subtle in individual scans, but strongly pointing-correlated, and therefore show up as large-scale features in the final maps. Additionally, they are present to varying extents in all feeds, have similar quantitative behavior in the map-domain, and can both be removed effectively with the map-PCA. The effects are discussed in the subsections below, with further analysis shown in Appendix C.

4.3.1. The turn-around effect

The so-called "turn-around" effect can be observed as strongly coherent excess power near the edges of the scan pattern, where the telescope reverses direction in azimuth. Illustrations of this effect can be seen in the first row, and partially in the third row, of Fig. 10. The feature manifests at the top and bottom of the maps, as this is where the telescope turn-around happens for Field 2 in equatorial coordinates. The feature oscillates slowly across the frequency domain, and the leading theory of its origin is some standing wave oscillation induced by mechanical vibrations. The effect is somewhat less pronounced in Season 2b, after the reduction in telescope pointing speed and acceleration, but it is still present.

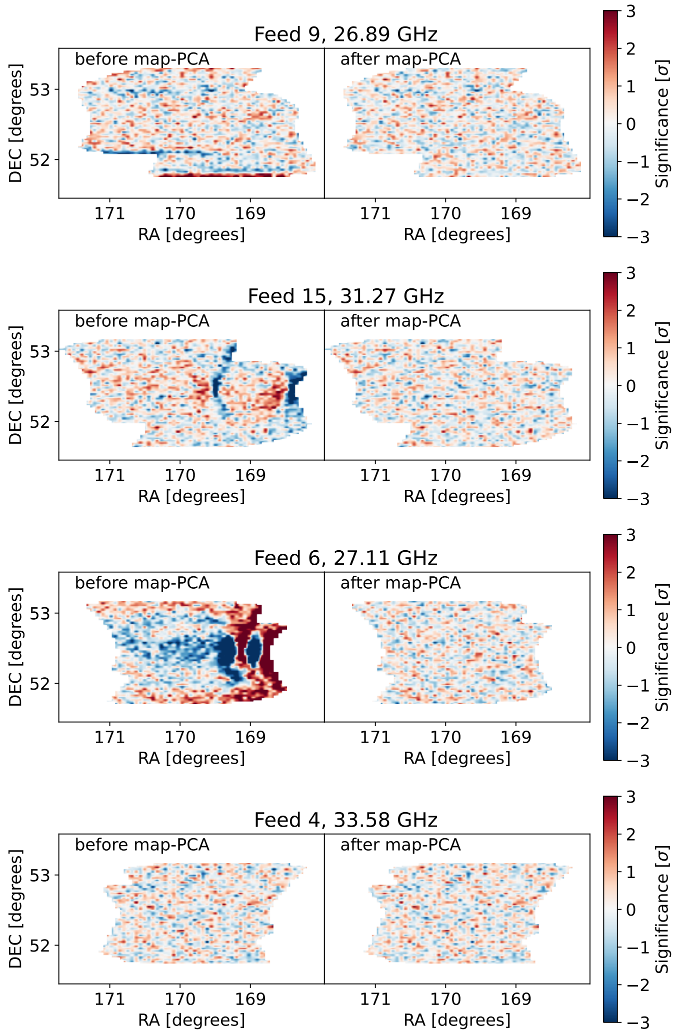


Fig. 10. Selection of individual frequency maps from Field 2 before *(left)* and after *(right)* the map-domain PCA filter. Because of the uneven sensitivity among the pixels, and to emphasize the relevant systematic effects, all maps have been divided by their white noise uncertainty. From top to bottom, each row shows maps that are dominated by (1) the turn-around effect; (2) the start-of-scan effect; (3) both effects simultaneously; and (4) neither effect. Both effects appear to manifest twice, on two slightly offset maps. This offset effect originates from the physical placement of the feeds, as they observe the fields as they are both rising and setting on the sky. In the equatorial coordinate system of these maps, the telescope scans vertically, and the field drifts from left to right.

4.3.2. The start-of-scan effect

A related effect is called the "start-of-scan" effect, which is a wave-like feature in frequency that occurs at the beginning of every scan and decays exponentially with a mean lifetime of around 19 sec. As the telescope always starts each scan at the same side of each sky field, this systematic effect shows up in the map domain as a strong feature on the Eastern edge of the map, as can be seen in the second and third rows of Fig. 10. Next to the strong positive or negative signal (this varies by frequency) at the very edge of the map, the opposite power, at lower amplitude, can be observed as we move Westward across the map. This opposite power is simply a ringing feature from the normalization performed during the pipeline (see Appendix C for details).

The exact origins of the "start-of-scan" effect are unknown, but the fact that it only happens at the beginning of scans (which are separated by a repointing to catch up with the field), and disappears during constant elevation scanning, suggests that a potential candidate is mechanical vibrations induced by the repointing. We also observe the effect to be mostly associated with one of the four DCM1s (first downconversion module), namely DCM1-2, relating to feeds 6, 14, 15, 16, and 17. The effect's strong correlation with DCM1-2 points to a possible source in the local oscillator cable, shared by the channels in a DCM1 module; imperfect isolation of the mixer would cause a weak common-mode resonance to manifest.

An important detail in this analysis is that, because of the normalization and $1/f$ filter in the TOD pipeline, any standing wave signal with a constant resonant cavity wavelength over time will be filtered away. For a standing wave to survive the filtering, it must have a changing wavelength. Prime suspects for the origin of this effect are therefore optical cavities that could expand or contract in size, or cables that could be stretched.

4.3.3. Effects of map-domain filtering

The map-domain PCA filtering was implemented in an attempt to mitigate these systematic effects and it has proved to be effective at this task. The first PCA component alone subtracts both effects to a level where they are not visible in the maps by eye. This shows that both effects are well modeled as the outer product of a pixel vector and a frequency vector.

Visually inspecting the PCA components, we usually see some structure for the first 3–5 modes. Figure 11 shows an example of the five leading PCA components and their amplitudes. Here we can see that the first component has very clear structure in both the map and frequency domain. The remaining modes seem to absorb some residuals after this first mode, especially on specific channels close to the edges of each of the two Bands that divide the frequency range in two.

We have chosen to remove just five out of 256 PCA components in the map-PCA filter, as no structure was visible by eye in the worst-affected cases after this number of components was removed, and the removal of more components did not significantly affect the results of any subsequent analysis (such as the power spectrum). With only five components being removed, we also limit the potential for CO signal loss. We could have employed a similar approach to the TOD PCA, with a dynamic amount of components, but the noise properties of the maps are more complicated than the TODs, and we have chosen to keep a static number of components, postponing more fine-tuning to future analysis. The filter is applied to the individual feed maps, and to individual splits – both the elevation split used for the cross power spectrum, and the individual map splits for the null tests.

With the application of the map-PCA filter, we observe that the start-of-scan and turn-around effects are suppressed well below the white noise of the maps, as can be seen in Fig. 10. We have also designed a null test to specifically target the turn-around systematic effect, by splitting the maps at the TOD level into east- and west-moving azimuthal directions. The turn-around effect manifests very differently in each half of this split, making it the basis for a sensitive null test. We find that the Season 2 data passes this test after the map-PCA filter has been applied (Stutzer et al. 2024).

The standard deviation of the maps only falls by 2% after applying the filter, as the noise still dominates the overall amplitude. However, smoothing the maps slightly will enhance large-scale correlations, while suppressing uncorrelated noise. Smoothing both the filtered and unfiltered 3D maps using a Gaussian with $\sigma = 3$ voxels, the standard deviation is 67% lower in the map-PCA filtered map. We can similarly observe that the average correlation between neighboring pixels (of the unsmoothed map), a good indication of the level of larger scale structure, falls from 6.3% to -0.4% after applying the map-PCA.

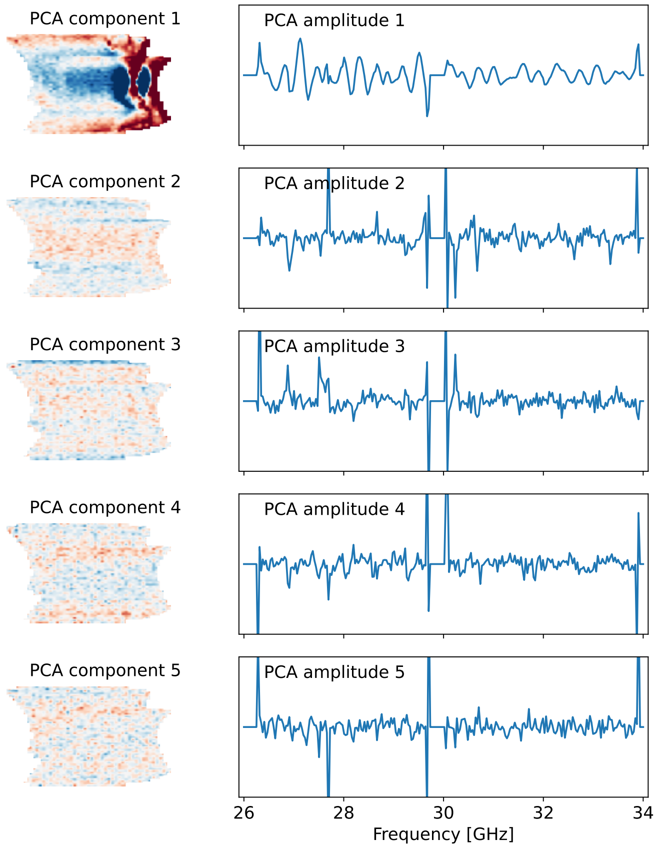


Fig. 11. Leading PCA components v^k *(left)* and their respective frequency amplitudes a^k *(right)* for Field 2 as observed by feed 6 prior to the map-PCA filter; this feed is the most sensitive to pointing-correlated systematics. All maps are divided by their respective uncertainties to highlight the key morphology. All rows share the same color range and y-axis scale, but the specific values have been omitted as they are not easily interpretable.

The magnitude of these systematic effects is different between feeds and frequencies, as we saw from Figure 10. Performing a χ^2 white noise consistency test on the individual frequency channel maps of each feed, we find that for the worst feeds, namely those associated with DCM1-2 (feeds 6, 14, 15, 16 and 17), around 50 % of their channels fail this test at $> 5\sigma$. The best-behaving feeds are 4, 5, 10, and 12, all with fewer than 5 % of channels failing at $> 3\sigma$. However, we want to emphasize both that no feed is completely without these effects before the map-PCA, and that after the map-PCA, there is no longer a quantitative difference between the "good" and "bad" feeds, with all feeds passing χ^2-tests at expected levels.

The PCA filtering (both in the time and map domain) constitutes the only non-linear processing in the pipeline. Non-linear filtering makes it more difficult to estimate the resulting signal bias and transfer function. In Sect. 6.4 we demonstrate that a PCA filter applied on a noisy matrix behaves linearly with respect to a very weak signal, and we find that any CO signal in the data is well within this safe regime.

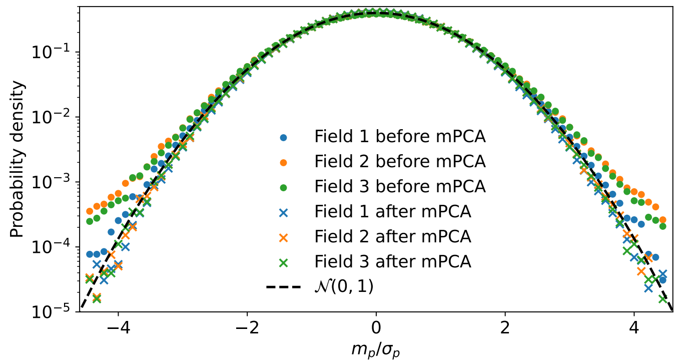


Fig. 12. Histogram of all map pixel temperature values across all feeds and frequencies, divided by their white noise uncertainty. The three fields are shown separately, before and after application of the map PCA filter. A normal distribution is shown in black; a completely white noise map will trace this distribution. All three fields show excess high-significance pixels before the map PCA. After the filter, all three fields fall slightly below the normal distribution on the wings, because of the slight over-subtraction of noise at various stages in the pipeline.

4.4. Final maps

Figure 12 shows the distribution of map voxel values for all three fields, in units of significance, before and after the map-PCA. Before the map-PCA, the distribution shows a clear excess on the tails, while the cleaned maps are very consistent with white noise. This is expected and desired, as the CO signal is so weak that individual frequency maps are still very much dominated by the system temperature. The noise level in the maps is, actually, about 2.5% lower than expected from the white noise uncertainty, due to the filtering in the pipeline. This effect can be seen in Fig. 12, with the histograms falling slightly below the normal distribution.

Figure 13 shows the distribution of voxel uncertainties over the three fields for this work and our ES maps. Each voxel has an approximate size of 2×2 arcmin, which, together with the frequency direction, corresponds to a comoving cosmological volume of $\sim 3.7 \times 3.7 \times 4.1\,\mathrm{Mpc}^3$. For Fields 2 and 3, the high sensitivity $< 50\,\mu$K region corresponds to a comoving cosmological cube of around $150 \times 150 \times 1000\,\mathrm{Mpc}^3$ per field. Combining all three fields, Season 2 has one million voxels with an uncertainty $< 50\,\mu$K, compared to one million voxels below $< 125\,\mu$K for Season 1. The footprint of the final maps have increased slightly in size because of the wider scan pattern of Season 2b.

The sensitivity increase per field over Season 1 is 2.0, 2.6, and 2.7, for Fields 1, 2, and 3, respectively. Fields 2 and 3 are now the highest sensitivity fields, while Field 1 is noticeably worse, from larger losses to data selection, especially Moon and Sun sidelobe pickup. The uncertainties are estimated from Eq. (11), and correspond well to the noise level observed in the map, as we saw from Fig. 12. A figure showing the uncertainties across the fields on the sky can be found in Appendix E, together with a subset of the final maps.

5. Data selection

In addition to a three-fold increase in observational hours, the second season also features a similar increase in data retention compared to the ES results. Table 3 compares the data loss in the ES results and this work. The table is split into three parts, namely 1) observational losses, 2) time- and map-domain losses, and 3) power spectrum domain losses. The Season 2 column

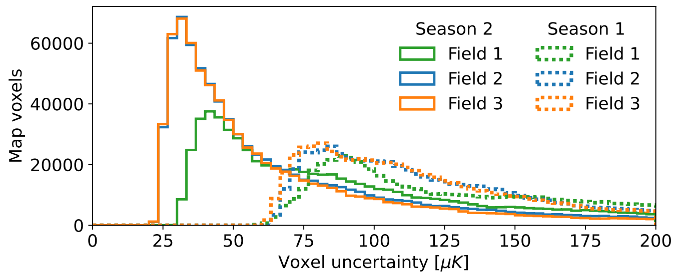


Fig. 13. Histograms comparing the map voxel uncertainties of this work (Season 2), and the ES publications (Season 1). The voxel values are for feed-coadded maps.

only relates to data taken during Season 2, but we have also reprocessed Season 1 data with the new pipeline for the final results.

5.1. Observational data retention

The first three rows of Table 3 are observational inefficiencies that have been corrected since the first season. E_{scan} constitutes the fraction of scans that were performed in constant elevation mode, as opposed to Lissajous scans which were cut due to large systematic effects (see Sect. 2.2). E_{feed} is the fraction of functioning feeds, and E_{el} is the fraction of data taken at elevation 35°–65° (see Sect. 2.1). Since Season 1, we no longer observe in Lissajous mode, all feeds are functional and the observing strategy has been optimized to maximize E_{el}. As a result, the total data retention from these three cuts, which was 32% in the first season, is now 100%.

5.2. Time and map domain data selection

In the next section in Table 3, E_{freq} refers to the frequency channel masking performed in `l2gen`, as discussed in Sect. 3.7. The masking algorithm itself is virtually identical to in ES, with a few changes. The shifting of aliased power into channels outside the nominal frequency range (Sect. 2.3) means that, from Season 2b onwards, we recover the 8% of channels masked in Season 1 and 2a. The inclusion of the new per-feed PCA filter in `l2gen` results in slightly fewer data being masked by data-driven tests. However, we have also increased the number of manually flagged channels that seem to be performing sub-optimally, leaving us with a E_{freq} data retention only slightly higher than for ES.

Next, E_{stats} constitutes the cuts performed in the `accept_mod` script, which discards scans based on different housekeeping data and summary statistics of the scans. There are over 50 such cuts in total, most of them removing a small number of outlier scans. Upon the completion of the Season 2 null test framework (Stutzer et al. 2024), null tests failed on five scan-level parameters. Cuts on these parameters were tightened or implemented in `accept_mod`, and the null tests now pass. The five new or tightened cuts are: 1) any rain during the scan; 2) wind speeds above 9 m/s; 3) high average amplitude of the fitted TOD PCA components; and 4–5) outliers in the f_{knee} of the 0th and 1st order $1/f$ filter components[8]. Additionally, all other `accept_mod` cuts from ES are continued,

[8] The $1/f$ filter fits the time-dependent components c_t^0 and c_t^1, primarily picking up correlated noise and changes in the atmosphere. We perform a $1/f$ fit to the components as functions of time, and cut when the f_{knee} falls outside the typical range of values

Table 3. Data retention overview.

	Season 1	Season 2	Explanation
E_{scan}	50.0%	100.0%	Retained scans (CESs)
E_{feed}	84.2%	100.0%	Functional feeds
E_{el}	75.6%	100.0%	Inside good elevation range
$\boldsymbol{E_{\text{obs}}}$	**31.8%**	**100.0%**	Observational data retention
E_{freq}	72.8%	74.3%	Frequency masking in `l2gen`.
E_{stats}	57.4%	33.6%	Cuts on accept-mod statistics
$E_{\chi^2_{P(k)}}$	72.2%	100.0%	Per-scan auto-PS χ^2-test
$\boldsymbol{E_{\text{cuts}}}$	**30.1%**	**24.9%**	Map-level data retention
$E_{\chi^2_{C(k)}}$	52.4%	100.0%	Cross-spectrum χ^2-test
$E_{C(k)}$	94.7%	75.0%	Cross-spectrum auto combinations
$\boldsymbol{E_{\text{PS}}}$	**49.6%**	**75.0%**	Retained data at PS-level
$\boldsymbol{S_{\text{tot}}}$	**6.8%**	**21.6%**	Final PS-domain sensitivity, calculated as $S_{\text{tot}} = \sqrt{E^2_{\text{obs}} E^2_{\text{cuts}} E_{\text{PS}}}$

Notes. Surviving fraction of data for different filtering steps of the pipeline. The left column shows the values used for the ES analysis, and the right column shows this work. The first 3 rows show individual data losses to observational constraints, which are combined in the gray row below. The three next rows show data retention to time and map domain cuts, again combined below. Finally, the next two rows show the losses in the power spectrum domain, also combined. The last row, S_{tot} shows the final fraction of theoretical power spectrum sensitivity from the combined data retention (see the text for details). The losses are multiplicative, such that multiplying E for all the individual losses gives the retained data fractions shown in gray.

and the surviving data fraction has therefore fallen noticeably, from 57.4 % to 33.6 %. No attempt has yet been made to tune these cuts, presenting us with future potential for increased data retention.

Finally, $E_{\chi^2_{P(k)}}$ is the last scan-level cut. Each scan is binned to a very low-resolution 3D map, and a series of χ^2-tests are performed on different 2D and 3D auto power spectra calculated from these maps. In the Season 2 pipeline, this cut is removed entirely, for two reasons. Firstly, we found little evidence that it helped us pass null tests or remove dangerous systematic errors from the final data. Secondly, we found it difficult to calculate robust pipeline transfer functions for each individual power spectrum, as individual scans might vary a lot in sky footprint and pointing pattern. We therefore saw little reason to keep this cut in the pipeline.

5.3. Power spectrum level data selection

The last section of Table 3 shows the fraction of data retained after cuts in the power spectrum domain; details on the power spectrum methodology are described by Stutzer et al. (2024). In summary, we calculate pseudo cross spectra between different groups of feeds and across pointing elevations and then average these spectra to get the CO power spectrum. However, some of the cross-spectra are discarded before averaging, and this is the loss discussed in this section. This loss in the power spectrum domain has to be tracked separately from data loss in the map and TOD domain, as the losses cannot naively be added together. Since map values are squared when calculating the power spectrum, so is the map-domain data volume when calculating the power spectrum sensitivity. The total power spectrum sensitivity is therefore calculated as $S_{\text{tot}} = \sqrt{E^2_{\text{obs}} E^2_{\text{cuts}} E_{\text{PS}}}$.

In this table section, $E_{\chi^2_{C(k)}}$ constitutes a χ^2-test on the individual feed-feed cross-spectra and cuts away any cross-spectrum with an average significance above 5σ. In ES, we lost around half the data to this cut. The cut has now been entirely removed, for several reasons. First of all, we now have a much more rigorous null test framework, and find that we pass all null tests without these cuts. Secondly, we strongly prefer moving all data-inferred cuts to a point as early in the pipeline as possible, to reduce any potential biasing effects. It is therefore a considerable pipeline improvement over ES that we now perform no data-inferred cuts in the power spectrum domain.

Finally, $E_{C(k)}$ is the fraction of the cross power spectra which are not auto-combinations between the same feeds. In ES we performed cross-spectra between all 19 feeds, which resulted in a loss of 19 out of 19×19 cross-spectra, or 5.3%. We now calculate cross-spectra between four groups of feeds, for better mitigation of systematic effects and improved overlap, resulting in a loss of 4 out of 4×4 feeds, or 25 %. This is a theoretical approximation of the sensitivity, as the varying degrees of overlap between different feeds will interplay with the sensitivity.

5.4. Future prospects for data selection

Combining the retained map-level data with the retained PS-level data we keep 21.6% of the theoretical sensitivity, compared to 6.8% for ES, a more than three-fold increase. Most of this increase comes from much higher observational data retention E_{obs}, and the removal of the $\chi^2_{P(k)}$ and $\chi^2_{C(k)}$ cuts. We now have no data-driven cuts in the power spectrum domain, where the signal is the strongest, leaving us less susceptible to signal bias.

In order to pass null tests and allow for the removal of other cuts, the data retention after cuts on `accept_mod` statistics, E_{stats}, has decreased quite substantially. We erred on the side of caution when introducing the new cuts, and once the data passed all the null tests we made no attempt at reclaiming any data from these cuts. In future analysis, we are therefore confident that better tuning of these parameters, assisted by an even better understanding and filtering of systematic effects, will allow us to substantially increase the amount of data retained at this step.

The numbers in Table 3 are averages across fields, feeds, and scans, and the combined data retentions, E_{map}, E_{PS} and S_{tot} are for simplicity calculated by naively multiplying together the individual retentions. This ignores certain complications, like correlations between the cuts, and the actual sensitivity might therefore differ slightly. It should also be noted that the right column constitutes the efficiency of Season 2 data in combination with the Season 2 pipeline, and re-analysis of Season 1 does not reach a S_{tot} of 21.6%, as the losses to E_{obs} will still apply even with the improvements to the pipeline.

6. Pipeline signal impact and updated transfer functions

The final maps are biased measurements of the CO signal, due to signal loss incurred in observation and data processing, leading to a biased power spectrum. This effect can be reversed by estimating a so-called transfer function $T(k_\parallel, k_\perp)$, which quantifies this signal loss at different scales. We separate the angular modes $k_\perp$ and the frequency/redshift modes $k_\parallel$, as the impact on the CO signal is usually very different in these two dimensions. In this section, we present updated versions of the three relevant transfer functions:

- The pipeline transfer function $T_p(k_\parallel, k_\perp)$: The time- and map-domain processing will inevitably remove some CO signal.
- The beam transfer function $T_b(k_\perp)$: The size and shape of the beam will suppress signal on smaller scales in the angular dimensions.
- The voxel window transfer function $T_v(k_\parallel, k_\perp)$: The finite resolution of the voxels suppresses signal on both angular and redshift scales close to the size of the voxels.

6.1. Updated Beam and Voxel Window Transfer functions

In the ES analysis, the beam and voxel window transfer functions were estimated using simulations (Ihle et al. 2022). However, because the voxel grid and beam of the COMAP instrument are well understood we can also compute $T_b(k_\perp)$ and $T_v(k_\parallel, k_\perp)$ analytically. As the COMAP mapmaker simply uses nearest neighbor binning of the TOD into equispaced voxels the map is smoothed by a $\mathrm{sinc}^2(x)$ function along each map axis. Specifically, the voxel window can be expressed as[9]

$$T_v(k_\parallel, k_\perp) = T_{\mathrm{freq}}(k_\parallel)T_{\mathrm{pix}}(k_\perp) = \mathrm{sinc}^2\left(\frac{\Delta x_\parallel k_\parallel}{2\pi}\right)\mathrm{sinc}^2\left(\frac{\Delta x_\perp k_\perp}{2\pi}\right), \tag{15}$$

where $\Delta x_\perp$ and $\Delta x_\parallel$ are the voxel sizes in angular and frequency directions. Specifically, we have voxel resolutions of $\Delta x_\perp \approx 3.7\,\mathrm{Mpc}$ and $\Delta x_\parallel \approx 4.1\,\mathrm{Mpc}$. Note that since the angular pixel window is approximately radially symmetric we have approximated $T_\perp(k_\perp) \approx T_{\mathrm{RA}}(k_{\mathrm{RA}}) \approx T_{\mathrm{Dec}}(k_{\mathrm{Dec}})$. Both the perpendicular and parallel voxel transfer functions can be seen in Figs. 14 and 15.

In principle, we could reduce the voxel window signal impact on smaller scales in both the angular and frequency dimensions by binning the maps into higher resolution voxels, shifting the decline of $T_{\mathrm{freq}}(k_\parallel)$ and $T_{\mathrm{pix}}(k_\perp)$ to higher k-values. In practice, however, the angular voxel window applies at a scale where the beam transfer function already suppresses the signal beyond recovery. Similarly, line broadening is expected to heavily attenuate the CO signal above $\sim 1\,\mathrm{Mpc}^{-1}$ (Chung et al. 2021), although the exact extent of line-broadening depends on galaxy properties that are not yet well constrained. Additionally, it would be more computationally costly to perform the analysis in higher resolution.

Next, given the radial beam profile $B(r)$ (see Fig. 2 of Ihle et al. 2022) and the convolution theorem we can obtain the beam transfer function as

$$T_b(k_\perp) = |\mathcal{F}\{B(r)\}|^2, \tag{16}$$

where r is the radius from the beam center, and $\mathcal{F}$ is the (2D) Fourier transform. As we assume the telescope beam to be radially symmetric the resulting beam smoothing transfer function will be a function of just $k_\perp = \sqrt{k_{\mathrm{RA}}^2 + k_{\mathrm{Dec}}^2}$ giving $T_b(k_\perp)$. The main-beam efficiency is taken into account in the same manner as Ihle et al. (2022) prior to computing the Fourier transform of the beam. As we can see in Figs. 14 and 15 the beam is by far the most dominant effect limiting our ability to recover the signal at smaller scales.

[9] Note that we use the convention where $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$.

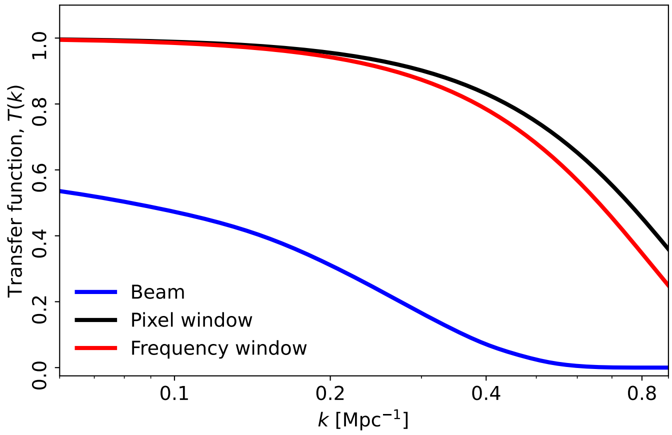


Fig. 14. Effective one-dimensional transfer functions for the instrumental beam (blue curve), pixel window (black curve), and frequency window (red curve) resulting from spherical averaging over the corresponding two-dimensional transfer functions shown in Fig. 15.

6.2. The signal injection pipeline

The last transfer function is that of the pipeline, which will inevitably remove some CO signal from the data. We estimate this impact by a signal injection pipeline, similar to what was done in ES (Foss et al. 2022).

We inject a simulated CO signal into the real Level 1 data before any filtering. The data are then propagated through the entire pipeline as usual, and the resulting mock observations are then compared to the known, unfiltered input signal to estimate the pipeline transfer function. We chose to inject the simulations into the actual data, instead of simulating the entire observation, in order to mimic the real systematic error and noise properties as closely as possible.

For the simulations we use approximate cosmological dark-matter-only simulations using the peak-patch method of Bond & Myers (1996) with updates by Stein et al. (2019). These are subsequently populated with CO emission using the COMAP fiducial model derived in Chung et al. (2022) ('UM+COLDz+COPSS'), which describes CO luminosities as a function of dark matter halo masses, $L_{\mathrm{CO}}(M_{\mathrm{halo}})$. These simulated mock maps are then boosted by a factor of 20, to recover a less noisy transfer function. This is counter-weighted by splitting the full dataset in 10 subsets, and passing each of these separately though the pipeline. We denote the maps as $\boldsymbol{s}^{\mathrm{mock}}_{\nu,\theta}$, where ν and θ are the frequency (redshift) and angular (pixel) dimensions, respectively. The TOD pipeline is quantitatively unaffected by the injection of this weak CO signal – even after a factor 20 boost, the brightest CO pixel in the simulation is still more than four orders of magnitude below the system temperature. In the map-domain, the SNR is much higher, and the implications of injecting a boosted signal will be discussed in Sect. 6.4, where we conclude that the map-PCA filter also behaves predictably for our chosen boost strength.

Using the real telescope pointing P, and estimated gain G and beam B, we get the signal-injected Level 1 data as

$$d^{\mathrm{mock}}_{t,\nu} = \mathsf{GPB}s^{\mathrm{mock}}_{\theta,\nu} + n_{t,\nu}, \tag{17}$$

where $n_{\nu,t}$ represents the actual Level 1 data, which acts as the noise term with respect to the injected mock CO signal. In order to mimic the observed CO signal as closely as possible, we also beam-smooth the maps used for the signal injection. The mock

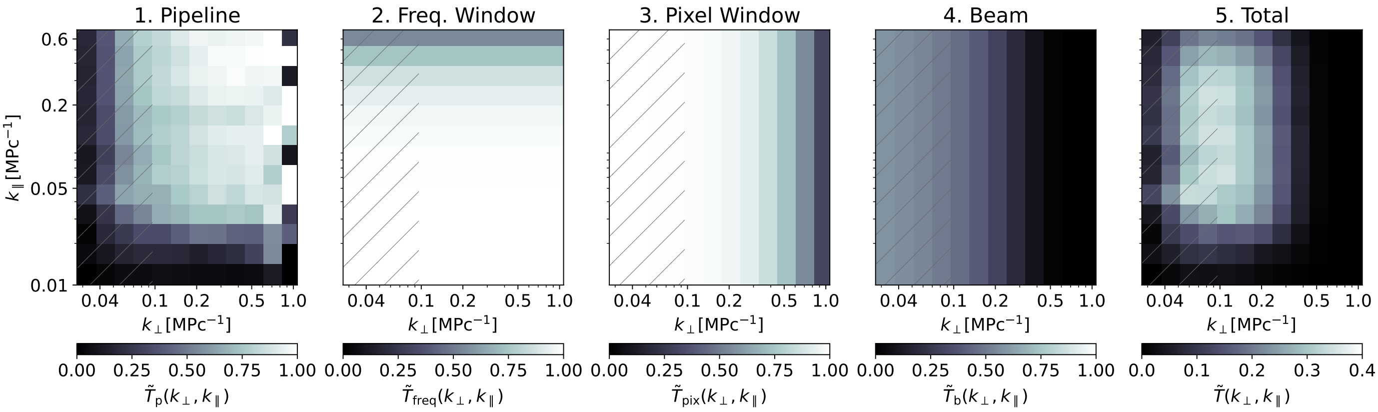


Fig. 15. Effective transfer functions used in the COMAP pipeline. From left to right, the five panels show 1) the filter transfer function, $T_\mathrm{f}(\boldsymbol{k})$, quantifying signal loss due to pipeline filters; 2) the pixel window transfer function $T_\mathrm{pix}(\boldsymbol{k})$ resulting from binning the TOD into a pixel grid; 3) the frequency window transfer function, $T_\mathrm{freq}(\boldsymbol{k})$, resulting from data down-sampling in frequency; 4) the beam smoothing transfer function, $T_\mathrm{b}(\boldsymbol{k})$; and 5) the full combined transfer function, $T(\boldsymbol{k})$, corresponding to the product of the four individual transfer functions. The striped region to the left is not used for our final analysis but is shown for completeness. Note that the leftmost panel has a colorbar that saturates at 0.4, unlike the other four.

data $d^\mathrm{mock}_{\nu,t}$ are then filtered by the pipeline to produce a mock map:

$$\begin{aligned} m^\mathrm{mock}_{\theta,\nu} &= f_\mathrm{map}\Big[M\Big(f_\mathrm{TOD}\Big[d^\mathrm{mock}_{\nu,t}\Big]\Big)\Big] \\ &= f_\mathrm{map}\Big[M\Big(f_\mathrm{TOD}\Big[\mathsf{GPB}s^\mathrm{mock}_{\nu,t} + n_{\nu,t}\Big]\Big)\Big], \end{aligned} \tag{18}$$

where we let f_TOD represent all time-domain filtering, M represents a noise-weighted binned map-maker, as described in Sect. 4.1, and f_map represents the map-domain PCA filter. Examples of resulting maps are shown in Appendix D.

To make sure that the reference CO simulation $\boldsymbol{s}^\mathrm{mock}_{\nu,\theta}$ is directly comparable to the filtered maps, we also perform some of the same treatment on it: we beam-smooth it, read it into a TOD with the real telescope pointing, and bin the TOD back into a map with the same resolution as the real maps. The difference, however, is that this is done completely without noise, and we do not apply any of the filters. We can write this as

$$\hat{s}^\mathrm{mock}_{\nu,\theta} = M\Big(\mathsf{GPB}s^\mathrm{mock}_{\nu,\theta}\Big). \tag{19}$$

Doing it this way means that we isolate the filter transfer function, and the effect of the beam and pixelation are not included. This is intentional, as we already estimated these impacts analytically in Sect. 6.1.

From these maps, we can now write the filter transfer function as

$$T = \frac{C(\boldsymbol{m}^\mathrm{mock}, \hat{\boldsymbol{s}}^\mathrm{mock})}{P(\hat{\boldsymbol{s}}^\mathrm{mock})}, \tag{20}$$

where the cross-spectrum, C in the numerator between the filtered mock data, $\boldsymbol{m}^\mathrm{mock}$, and the unfiltered mock signal, $\hat{\boldsymbol{s}}^\mathrm{mock}$, picks up all common signal modes after filtering while canceling residual systematic effects and noise in the mock data. The cross-spectrum is divided by the unfiltered signal auto spectrum, $P_{\boldsymbol{k}}(\hat{\boldsymbol{s}}^\mathrm{mock})$, to obtain a filter transfer function T.

Equation (20) represents a more robust estimator of the transfer function than the one employed in ES (Foss et al. (2022) Eq. 34), both because it does not require an accurate estimation of the noise power spectrum, and because using the cross-spectrum estimator as opposed to the auto-spectrum estimator makes it less susceptible to picking up signal in the data which does not originate from the injected CO, e.g. from systematic effects.

For the signal injection, we use all scans from Seasons 1 and 2a for Field 2. Preliminary analysis of the transfer functions of Fields 1 and 3, and the slower pointing scans of Season 2b, show that they are very similar, especially in the k-regime included in this work. As mentioned in the beginning of the section, we divide the scans into 10 random and equally large parts. Different dark matter halo simulations are injected into each part. This both reduces the impact of sample variance in the simulations and allows us to boost the signal a bit more without having to worry about PCA non-linearity (see the next section). We then average over the 10 resulting transfer functions, to get the final transfer function estimate.

6.3. Updated filter transfer function

The left-most panel of Fig. 15 shows the full COMAP pipeline transfer function, as described in the previous section, in parallel and perpendicular directions (i.e., frequency/redshift and angular scale, respectively). The Season 2 publications (Stutzer et al. 2024; Chung et al. 2024) exclude some of the larger angular scales accessible in the maps due to concerns about mode mixing and unconstrained modes due to poor overlap. For reference, the cutoff value for $k_\perp$ at $0.93\,\mathrm{MPc}^{-1}$ corresponds to angular scales of 36.4 arcmin.

Figure 16 shows the individual contributions of each filter to the full transfer function. We are somewhat limited on large angular scales by the normalization and pointing filters, and very limited on large redshift scales by the $1/f$ filter. The right-most column is noisy because the beam suppresses most of the signal at these small scales.

We also note that a small issue was discovered with the transfer function analysis published in Ihle et al. (2022), relating to how the mock signal was interpolated when injected into the TOD. The effect was that the transfer functions from ES was slightly underestimated, and the transfer function from Ihle et al. (2022) peaked at around 0.8. This issue has now been solved, and the new transfer function peaks correctly at almost 1.0.

6.4. Linearity of PCA filtering

All filters except the various PCAs constitute linear operations on the data. Linearity makes transfer function estimation much

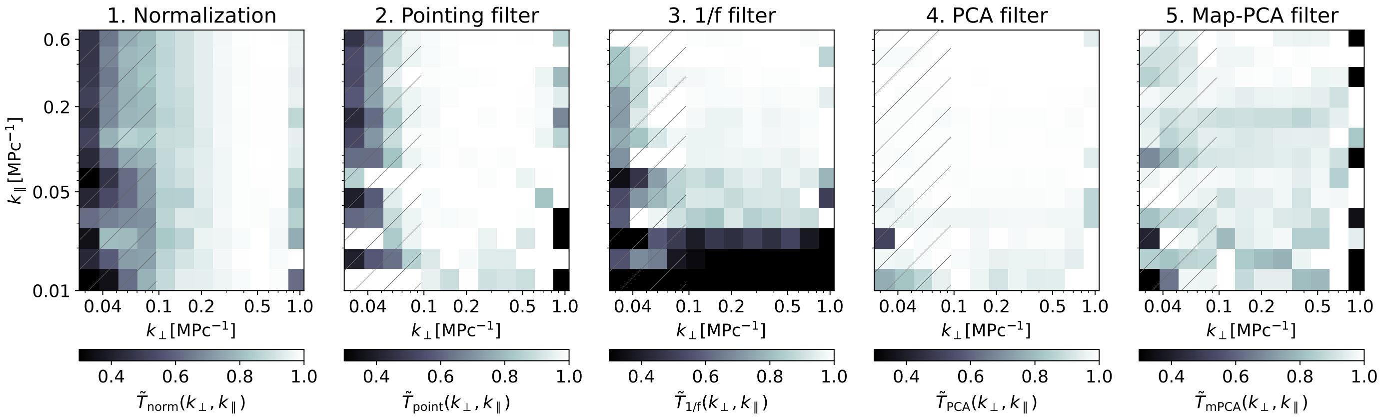


Fig. 16. Transfer functions for each of the five individual filters used in the pipeline. The normalization and pointing filters suppress large angular scales, while the $1/f$ filter almost entirely eliminates parallel modes larger than $k = 0.02\,\mathrm{Mpc}^{-1}$. The TOD PCA filter has almost no impact on the signal, primarily because so few modes are subtracted. The map PCA has a more noticeable signal loss, but it remains relatively scale-independent because the signal is still weak enough (see Sect. 6.4). The total pipeline transfer function shown in the leftmost panel of Fig. 15 is the product of these. The striped region to the left is not used in Season 2 results.

simpler, as neither the choice of CO signal model nor its level with respect to the noise impacts the resulting transfer function. For the PCA filters, both these factors could in principle impact the shape of the transfer function.

To quantify the sensitivity of the transfer function to such factors, we constructed a simplified version of the signal-injection pipeline. In order to be able to run many simulations, we made the following alterations to the pipeline:

– We bypass the time domain, and perform the signal injection in the map domain. This is the domain where the CO SNR is the strongest, and the map-domain processing is much more computationally efficient than that for the time-domain. The map mocks are the same as in Sect. 6.2, with some boost factor b.
– Instead of the real map, we use white noise simulations, drawn from the white noise uncertainty of Field 2.
– The resulting transfer function is calculated using Eq. (20), and then averaged across the simulations and feeds.

This process is repeated for 40,000 noise realizations for each of the 19 feeds, with boosts between 0.3 and 300 relative to the fiducial CO model of Chung et al. (2022). The variation of the resulting average transfer functions with boost strength can be seen in Fig. 17. The figure shows two distinct regimes. In the low SNR regime to around SNR=0.02 (boost 10), the PCA filter behaves linearly *with respect to the CO signal*. This is demonstrated by the independence of the transfer function to the SNR in this regime. Additionally, all the k-points lie on top of each other around a value of 0.96, meaning that all scales are suppressed at the same level because the PCA is simply fitting and subtracting random white noise. In the second regime, at high SNR, the transfer function is strongly scale-dependent but flattens out as the signal dominates the noise.

In conclusion, for a noisy matrix with an accompanying signal, a PCA filter behaves linearly with respect to the signal for a sufficiently weak signal. For us this means that when estimating the transfer functions, we need to use a sufficiently low boost value to avoid biasing our estimate of the transfer function. The individual data chunks used to estimate the transfer function are well within this linear regime, at an SNR of 0.004. The actual CO signal is of course of unknown amplitude, but assuming the fiducial model of Chung et al. (2022) results in an SNR of 0.002. If the CO signal were even close to the unsafe regime of SNR > 0.02, we would already have made a strong detection of it, as this would correspond to a 100 times brighter power spectrum than the fiducial model. We also note that in the future, as the experiment's sensitivity increases, we can simply perform the map PCA on sub-divisions of the data to keep the SNR low, as we already do on Season 1 + 2a and Season 2b, due to their differing pointing strategy.

This analysis was performed on the map-domain PCA filter, as the map domain is where the CO SNR is the strongest. Equivalent analysis has been performed for the TOD PCA filter, but the CO signal is so weak at the per-scan level that a boost factor of 2000 or greater is required to make it behave non-linearly with respect to the CO signal. The final transfer function is currently estimated jointly for all filters, but in future work we intend to estimate the transfer function for the mPCA separately from the other filters, gaining higher sensitivity on the linear parts of the pipeline that do not require as low a boost.

7. Summary and Conclusions

We have presented the improvements in data analysis, filtering, and data selection which have enabled us to increase the power spectrum sensitivity to 21.6% of the theoretical maximum, up from 6.8% in the ES publications. Combined with an increased integration time, the two most sensitive Season 2 fields both have a voxel uncertainty of $< 50\,\mu$K across $\sim 1.5° \times 1.5°$ patches on the sky. Across the three fields, this corresponds to a 2.5 times decrease in the total map uncertainty.

The largest increase in data retention comes from improvements in observational strategy. We now solely observe using CES, whereas in Season 1 50% of observations used Lissajous scans which proved prone to systematic effects and were not included in our ES analysis. Additionally, we now observe within elevation boundaries of 35°–65°, a region with minimal gradients in ground sidelobe pickup. In Season 1, 25% of scans fell outside this range and were discarded.

Additional increases in data retention have come from the removal of data cuts. In Season 1, power spectrum χ^2 tests were performed both on individual scan maps and on individual feed-feed cross-spectra before averaging them. These cuts removed (respectively) 28% of scans and 48% of cross-spectra, and are

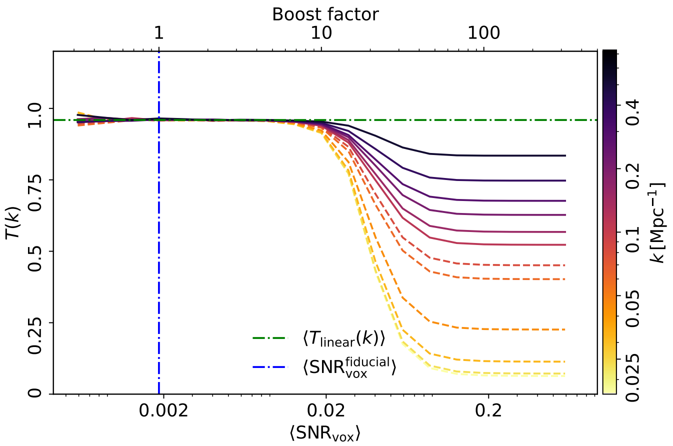


Fig. 17. Map PCA transfer function $T(k)$ as a function of the voxel SNR, with the different k-scales shown as differently colored lines. The equivalent boost used to the fiducial CO model is shown as the top x-axis. Modes with k-values that fall within the analysis bounds are shown as solid, while modes outside the scope of COMAP Season 2 are shown as striped. The horizontal green line shows the average transfer function in the low-SNR regime, which is 0.96. The vertical blue line shows the SNR that the maps would have if the fiducial model of Chung et al. (2022) perfectly described the true CO signal.

now no longer applied. This also reduces the possibility of signal bias owing to data-inferred cuts late in the pipeline.

In order to pass null tests and allow for the removal of other cuts, scan-level data cuts were implemented on six new housekeeping statistics, increasing the data lost to such cuts from 43% to 66%. However, no attempt was made to reclaim data after the null tests had passed, and the necessity of the cuts carried over from the ES pipeline is largely untested. We are confident this number can be greatly reduced in future work.

The removal of data cuts was also made possible by better mitigation of systematic errors. The time-domain pipeline has numerous smaller improvements, such as a new per-feed PCA filter, dynamically determined number of PCA components, better masking of the T_{sys} spikes, and more manual masking of consistently problematic channels. Most impactful, however, was the introduction of a map-level PCA filter, which proved essential to dealing with a couple of pointing-correlated systematic errors that emerged due to increased sensitivity. We show that the map PCA filter suppresses these effects to below the noise level, and decreases the standard deviation of slightly smoothed maps by 67%, to a level consistent with the expected white noise.

Although the PCA filters constitute non-linear filtering, we have shown that the PCA filters behave linearly with respect to any sufficiently weak signal. We find that the expected CO signal falls well within this regime, substantially simplifying transfer function estimation. We repeat the signal-injection pipeline transfer function estimation of the ES publications and ensure that the injected signal is also weak enough to maintain the PCA filters in their linear range. We have also replaced the simulation-based estimates of the beam and voxel windows transfer functions with more robust analytic expressions, improving their reliability at small scales.

COMAP has thus greatly increased its integration speed both through observational improvements, better processing, and reduced data cuts. Our final sensitivity retention of 21.6% of the theoretical maximum still leaves significant room for improvement, and we aim to increase this further as the Pathfinder continues to observe.

Acknowledgements. We acknowledge support from the Research Council of Norway through grants 251328 and 274990, and from the European Research Council (ERC) under the Horizon 2020 Research and Innovation Program (Grant agreement No. 772253 bits2cosmology and 819478 Cosmoglobe) This material is based upon work supported by the National Science Foundation under Grant Nos. 1517108, 1517288, 1517598, 1518282, 1910999, and 2206834, as well as by the Keck Institute for Space Studies under "The First Billion Years: A Technical Development Program for Spectral Line Observations". Parts of the work were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. HP acknowledges support from the Swiss National Science Foundation via Ambizione Grant PZ00P2_179934. SEH and CD acknowledge funding from an STFC Consolidated Grant (ST/P000649/1) and a UKSA grant (ST/Y005945/1) funding LiteBIRD foreground activities. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2024-00340759). PCB was supported by the James Arthur Postdoctoral Fellowship. DTC was supported by a CITA/Dunlap Institute postdoctoral fellowship for much of this work. The Dunlap Institute is funded through an endowment established by the David Dunlap family and the University of Toronto. Research in Canada is supported by NSERC and CIFAR. JGSL and NOS extend a great thanks to Sigurd K. Næss for all the fruitful discussions in the office, and while biking through nature, during the last three years. This work was first presented at the Line Intensity Mapping 2024 conference held in Urbana, Illinois; we thank Joaquin Vieira and the other organizers for their hospitality and the participants for useful discussions.

Software acknowledgments. Matplotlib for plotting (Hunter 2007); NumPy (Harris et al. 2020) and SciPy (Virtanen et al. 2020) for efficient numerics and array handling in Python; Astropy a community made core Python package for astronomy (Astropy Collaboration et al. 2013, 2018, 2022); Multi-node parallelization with MPI for Python (Dalcín et al. 2005, 2008; Dalcin et al. 2011; Dalcin & Fang 2021); Pixell (`https://github.com/simonsobs/pixell`) for handling sky maps in rectangular pixelization.

References

Akeson, R., Armus, L., Bachelet, E., et al. 2019, The Wide Field Infrared Survey Telescope: 100 Hubbles for the 2020s
Anderson, C. J., Luciw, N. J., Li, Y. C., et al. 2018, Monthly Notices of the Royal Astronomical Society, 476, 3382–3392
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bandura, K., Addison, G. E., Amiri, M., et al. 2014, in Ground-based and Airborne Telescopes V, ed. L. M. Stepp, R. Gilmozzi, & H. J. Hall (SPIE)
Bernal, J. L. & Kovetz, E. D. 2022, The Astronomy and Astrophysics Review, 30
Bond, J. R. & Myers, S. T. 1996, ApJS, 103, 1
Breysse, P. C., Chung, D. T., Cleary, K. A., et al. 2022, The Astrophysical Journal, 933, 188
Breysse, P. C., Kovetz, E. D., & Kamionkowski, M. 2014, MNRAS, 443, 3506
CCAT-Prime Collaboration, Aravena, M., Austermann, J. E., et al. 2022, The Astrophysical Journal Supplement Series, 264, 7
Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Nature, 466, 463
Chung, D. T., Breysse, P. C., Cleary, K. A., et al. 2024, arXiv e-prints, arXiv:2406.07512
Chung, D. T., Breysse, P. C., Cleary, K. A., et al. 2022, ApJ, 933, 186
Chung, D. T., Breysse, P. C., Ihle, H. T., et al. 2021, ApJ, 923, 188
Cleary, K. A., Borowska, J., Breysse, P. C., et al. 2022, The Astrophysical Journal, 933, 182
Crites, A., Bock, J., Bradford, C., et al. 2014, in Millimeter, submillimeter, and far-infrared detectors and instrumentation for astronomy VII, Vol. 9153, SPIE, 613–621
Dalcin, L. & Fang, Y.-L. L. 2021, Computing in Science & Engineering, 23, 47
Dalcin, L. D., Paz, R. R., Kler, P. A., & Cosimo, A. 2011, Advances in Water Resources, 34, 1124
Dalcín, L., Paz, R., & Storti, M. 2005, Journal of Parallel and Distributed Computing, 65, 1108
Dalcín, L., Paz, R., Storti, M., & D'Elía, J. 2008, Journal of Parallel and Distributed Computing, 68, 655
Dunne, D. A., Cleary, K. A., Breysse, P. C., et al. 2024, COMAP Early Science: VIII. A Joint Stacking Analysis with eBOSS Quasars
Fasano, A., Ade, P., Aravena, M., et al. 2024, in European Physical Journal Web of Conferences, Vol. 293, European Physical Journal Web of Conferences, 00018

Foss, M. K., Ihle, H. T., Borowska, J., et al. 2022, The Astrophysical Journal, 933, 184
Gabriel, K. R. & Zamir, S. 1979, Technometrics, 21, 489
Gailly, J.-l. & Adler, M. 1992, GNU Operating System
Geman, S. 1980, The Annals of Probability, 8, 252
Halko, N., Martinsson, P. G., & Tropp, J. A. 2011, SIAM Review, 53, 217
Harper, S. E., Dickinson, C., Cleary, K. A., et al. 2024, COMAP Galactic Science I: Observations of Spinning Dust Emission at 30GHz in Dark Clouds Surrounding the λ-Orionis Hii Region
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90
Ihle, H. T., Borowska, J., Cleary, K. A., et al. 2022, The Astrophysical Journal, 933, 185
Karkare, K. S., Anderson, A. J., Barry, P. S., et al. 2022, Journal of Low Temperature Physics, 209, 758–765
Keihänen, E., Keskitalo, R., Kurki-Suonio, H., Poutanen, T., & Sirviö, A.-S. 2010, Astronomy and Astrophysics, 510, A57
Koranne, S. 2011, in Handbook of Open Source Tools (Springer), 191–200
Korngut, P. M., Bock, J. J., Akeson, R., et al. 2018, in Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, Vol. 10698, SPIE, 576–586
Kovetz, E., Breysse, P. C., Lidz, A., et al. 2019, BAAS, 51, 101
Kovetz, E. D., Viero, M. P., Lidz, A., et al. 2017, arXiv e-prints, arXiv:1709.09066
Lamb, J. W., Cleary, K. A., Woody, D. P., et al. 2022, The Astrophysical Journal, 933, 183
Li, T. Y., Wechsler, R. H., Devaraj, K., & Church, S. E. 2016, The Astrophysical Journal, 817, 169
Lidz, A., Furlanetto, S. R., Oh, S. P., et al. 2011, ApJ, 741, 70
Loeb, A. & Zaldarriaga, M. 2004, Physical Review Letters, 92
Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, The Astrophysical Journal, 763, L20
Pearson, K. 1901, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2, 559
Pullen, A. R., Breysse, P. C., Oxholm, T., et al. 2023, Monthly Notices of the Royal Astronomical Society, 521, 6124–6142
Pullen, A. R., Chang, T.-C., Doré, O., & Lidz, A. 2013, ApJ, 768, 15
Rennie, T. J., Harper, S. E., Dickinson, C., et al. 2022, The Astrophysical Journal, 933, 187
Rudelson, M. & Vershynin, R. 2010, in Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures, World Scientific, 1576–1602
Santos, M. G., Cluver, M., Hilton, M., et al. 2017, MeerKLASS: MeerKAT Large Area Synoptic Survey
Stein, G., Alvarez, M. A., & Bond, J. R. 2019, MNRAS, 483, 2236
Stutzer, N. O., Lunde, J. G. S., Breysse, P. C., et al. 2024, arXiv e-prints, arXiv:2406.07511
Tamuz, O., Mazeh, T., & Zucker, S. 2005, Monthly Notices of the Royal Astronomical Society, 356, 1466
Tegmark, M. 1997, Physical Review D, 56, 4514–4529
Vershynin, R. 2010, arXiv preprint arXiv:1011.3027
Vieira, J., Aguirre, J., Bradford, C. M., et al. 2020, The Terahertz Intensity Mapper (TIM): a Next-Generation Experiment for Galaxy Evolution Studies
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261

Appendix A: Dynamic PCA threshold

It is known that the largest singular value of a Gaussian $P \times N$ matrix with variance σ^2 can be approximated as $\lambda \approx C(P,N) \cdot \sigma(\sqrt{P} + \sqrt{N})$ for large matrices (Geman 1980; Rudelson & Vershynin 2010; Vershynin 2010), where $C(P,N) \approx 1$ is some correction-factor for which no reliable theoretical model exists. We, therefore, simulated 50,000 noise matrices and empirically solved for $C(P,N)$ within the relevant regime of N from 5,000 to 30,000, and P from 100 to 80,000, which captures all matrix sizes our pipeline will encounter. We find that the correction factor can, in our size range, be well-modeled as

$$C(P,N) = 1.00476 - 0.00396 \cdot \log(P) \cdot \log(N) + 0.0000876 \cdot \log(P)^2 \cdot \log(N)^2. \tag{A.1}$$

We have no doubt that this factor is unlikely to extrapolate sensibly beyond the range we tested it in, but that is of no concern. The relative error to the mean of our 50,000 simulations is less than 0.1% within our defined bounds.

Appendix B: The map-domain PCA filter

Appendix B.1: The effect of noise-weighting on PCA

This section expands on the PCA discussion of Sect. 4.2, and we keep the same variable and index names, for easier comparison. The first principal component $\boldsymbol{w}^1$ and its amplitude $\boldsymbol{a}^1$ of a PCA[10] for a data-matrix $\boldsymbol{m}$ are the vectors that minimize

$$f(\boldsymbol{a},\boldsymbol{v}) = \sum_\nu \sum_p (m_{\nu,p} - a_\nu w_p)^2. \tag{B.1}$$

We might, however, want to find these vectors while weighting the elements of $\boldsymbol{m}$, for example, if it is non-uniformly noisy. If the weights themselves can be separated into an outer product, such that we have row and column uncertainties $\boldsymbol{\sigma}^{\rm row}$ and $\boldsymbol{\sigma}^{\rm col}$, this can trivially be done by inverse-variance weighting the matrix elements of the expression above. We now minimize

$$f(\boldsymbol{a},\boldsymbol{w}) = \sum_\nu \sum_p \frac{(m_{\nu,p} - a_\nu w_p)^2}{\sigma_\nu^{\rm row}\sigma_p^{\rm col}}, \tag{B.2}$$

which can be expanded to

$$f(\boldsymbol{a},\boldsymbol{w}) = \sum_\nu \sum_p \left(\frac{m_{\nu,p}}{\sqrt{\sigma_\nu^{\rm row}\sigma_p^{\rm col}}} - \frac{a_\nu}{\sqrt{\sigma_\nu^{\rm row}}} \frac{v_p}{\sqrt{\sigma_p^{\rm col}}} \right)^2. \tag{B.3}$$

This is still a valid PCA problem, on the same form as Eq. (B.1), with $\boldsymbol{a}' = \boldsymbol{a}/\boldsymbol{\sigma}^{\rm row}$ and $\boldsymbol{w}' = \boldsymbol{w}/\boldsymbol{\sigma}^{\rm col}$ now being the amplitude and component we fit for. If $\boldsymbol{m}$ contains a feature that can be decomposed into an outer product, this will be recovered by $\boldsymbol{w} = \boldsymbol{w}'\boldsymbol{\sigma}^{\rm row}$ and $\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{\sigma}^{\rm col}$.

However, if we want to weight every element of $\boldsymbol{m}$ with arbitrary uncertainty $\sigma_{\nu,p}$, this no longer holds. We can still write the problem simply as

$$f(\boldsymbol{a},\boldsymbol{w}) = \sum_\nu \sum_p \frac{(m_{\nu,p} - a_\nu w_p)^2}{\sigma_{\nu,p}^2} \tag{B.4}$$

which can be expanded to

$$f(\boldsymbol{a},\boldsymbol{w}) = \sum_\nu \sum_p \left(\frac{m_{\nu,p}}{\sigma_{\nu,p}} - \frac{a_\nu}{\sqrt{\sigma_{\nu,p}}} \frac{w_p}{\sqrt{\sigma_{\nu,p}}} \right)^2, \tag{B.5}$$

but because $a_\nu/\sigma_{\nu,p}$ and $w_p/\sigma_{\nu,p}$ are now matrices and not vectors, this minimization problem is no longer a PCA. We therefore have no way of recovering the desired $\boldsymbol{a}$ and $\boldsymbol{w}$ corresponding to the matrix $\boldsymbol{m}$ if we perform the regular PCA algorithm on the matrix $m_{\nu,p}/\sigma_{\nu,p}$, for a general $\sigma_{\nu,p}$. We must therefore find a different way of minimizing B.4, which is discussed in the following section.

Appendix B.2: Generalization of the PCA algorithm

With the generalization of Eq. (B.4), we can no longer utilize the usual methods of solving a PCA problem, such as the SVD. Instead we employ the technique suggested by Tamuz et al. (2005) and Gabriel & Zamir (1979), where we iteratively make improved guesses at $\boldsymbol{w}$ and $\boldsymbol{a}$:

1. Make an initial guess at $\boldsymbol{a}$ and $\boldsymbol{w}$, either completely random or informed by some knowledge of the data.
2. Solve for the optimal $\boldsymbol{a}$ while holding the current $\boldsymbol{w}$ constant by differentiating Eq. (B.4), holding $\frac{{\rm d}f(\boldsymbol{a},\boldsymbol{w})}{{\rm d}\boldsymbol{a}} = 0$, and solving for $\boldsymbol{a}$ as
$$a_\nu = \frac{\sum_p \frac{m_{\nu,p} w_p}{\sigma_{\nu,p}^2}}{\sum_p \frac{w_p^2}{\sigma_{\nu,p}^2}}. \tag{B.6}$$
3. Given the new $\boldsymbol{a}$, calculate $\frac{{\rm d}f(\boldsymbol{a},\boldsymbol{w})}{{\rm d}\boldsymbol{w}} = 0$, and solve for the new optimal w_p as
$$w_p = \frac{\sum_\nu \frac{m_{\nu,p} a_\nu}{\sigma_{\nu,p}^2}}{\sum_\nu \frac{a_\nu^2}{\sigma_{\nu,p}^2}}. \tag{B.7}$$
4. Repeat 2. and 3. until the incremental changes in $\boldsymbol{a}$ and $\boldsymbol{w}$ are below some chosen threshold ϵ.

Although we cannot prove that this is a convex problem and that the optimal solution is guaranteed, we have never seen it converge to an unreasonable solution. Additional robustness can be achieved by repeating the fit with different initial guesses, and confirming that they converge to the same solution.

The algorithm will converge on the same solution as the regular PCA in the case of uniform weights $\sigma_{\nu,p} = 1$, or where the weights can be perfectly decomposed into an outer product of weights in rows and columns, as in Eq. (B.2).

Appendix B.2.1: Multiple components

Using this decomposition, we can fit multiple components, similar to the ordered set of principal components in the PCA, by simply subtracting the previous components from the data. We then simply define $a_i^{(1)}$ and $m_j^{(1)}$ as the results from the previous section, and let

$$m_{\nu,p}^{(1)} = m_{\nu,p} - a_\nu^1 w_p^1. \tag{B.8}$$

We can then find $a_\nu^{(2)}$ and $w_p^{(2)}$ by performing the procedure from the previous section on $m_{\nu,p}^{(1)}$. This is again, in the case of $\sigma_{\nu,p} = 1$, entirely equivalent to finding the largest principal components of $m_{\nu,p}$.

[10] In the regular PCA formalism, the eigenvectors $\boldsymbol{w}$ are typically unit vectors of length 1, which is not automatically the case throughout this section. However, $\boldsymbol{w}$ can be normalized to 1 at any point by inversely adjusting $\boldsymbol{a}$. In a regular PCA, a (scalar) singular value is present in the solution. In the formalism presented here, there is no explicit singular value, and it can be absorbed into the amplitudes $\boldsymbol{a}$.

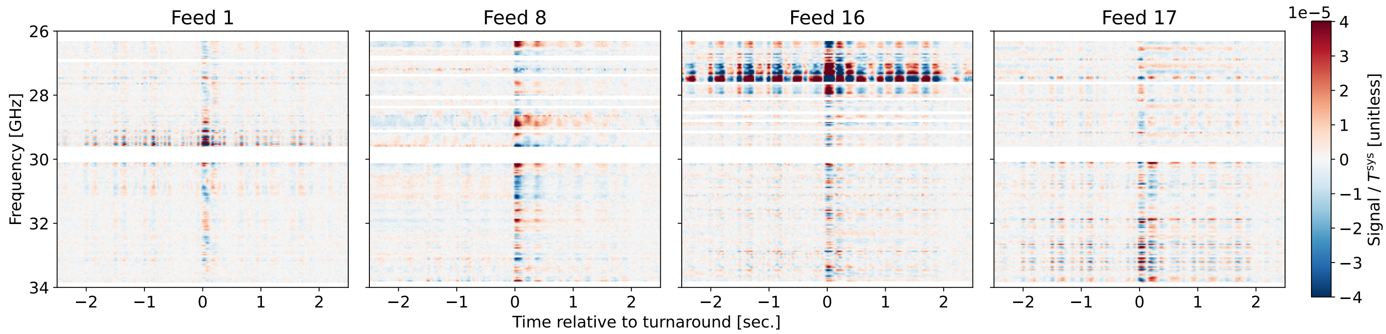


Fig. C.1. Filtered time-domain data stacked on the turn-around of the telescope, to emphasize the turn-around systematic effect, for four selected feeds. The data is an average across all turn-arounds of all available Level 2 scans. The data is divided by the system temperature of the data in each channel, and the values therefore represent the signal strength of the effect relative to the average white noise level in the scans.

Appendix C: Start-of-scan and turn-around effects

As outlined in Sect. 4.3, our increased sensitivity has revealed two systematic effects in our maps which were not discovered in our ES publications. We here explore these effects in more detail, especially in the time-domain, where these effects are easier to understand than in the map-domain. Note that the data shown in this section has not been filtered by the map-PCA filter, and the systematic effects demonstrated are (to the best of our analysis) not present in our final maps.

Appendix C.1: The turn-around effect

The turn-around effect is a sharp feature located around the azimuth edges of the scanning pattern, where the telescope turns. In Fields 2 and 3, which rise and set almost vertically across the sky (see Figure 1) and are therefore observed at almost the same angle at all times, this effect manifests as sharp edges on the top and bottom (i.e. the highest/lowest declinations) of the equatorial coordinate maps. This can be seen in the first and third rows of Fig. 10.

To better understand this systematic effect, we have extracted the data around the telescope turn-arounds for all our scans, and stacked the result on the turn-around time. Figure C.1 shows the result for four selected feeds. For all four feeds, we see a feature that peaks around the turn but is also present both leading up to and after the turn. In the frequency direction the feature has a slow wave-like feature. The feature manifests differently in different feeds and frequencies but has in common that it is wave-like both in frequency and time and peaks in power around the turn-around. As the telescope turn-arounds are the regions with the highest acceleration, a likely origin of this effect is some standing wave induced by the mechanical vibrations of the azimuth drive. Some feeds show significantly stronger manifestations of the turn-around effect than others, but all feeds are affected to some extent, and no explanation has yet been found as to why feeds are affected differently.

Attempts have been made to model this effect in the time-domain. This has proven difficult, among other reasons because the effect is actually very weak compared to the noise level in a single scan. As seen from Fig. C.1, the effect peaks at more than four orders of magnitude below the noise temperature of the telescope. The effect is only visible in the final maps because it seems strongly coherent across different scans. However, because we have to combine thousands of scans in order to observe the effect, it is also difficult to assess if the effect is indeed perfectly coherent across all scans, or if we are simply observing the average impact of this effect. The efforts of modeling the effect in the time domain was also made somewhat moot by how effective the map-PCA was at removing the effect in the map-domain.

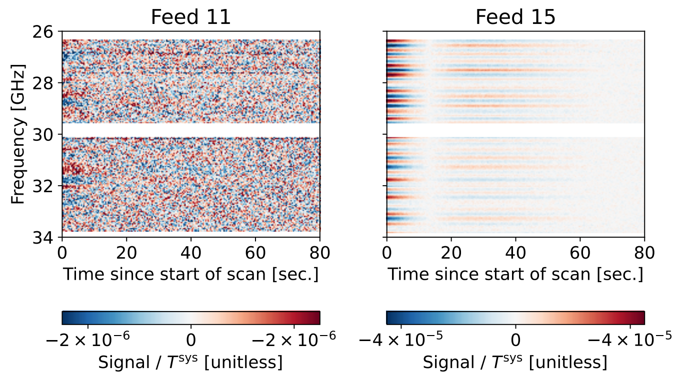


Fig. C.2. Same setup as Figure C.1, but stacked on the beginning of each scan, to emphasize the start-of-scan systematic effect. Notice the difference in the colorbar limits.

Appendix C.2: The start-of-scan effect

The start-of-scan effect is similar to the turn-around effect in that it is also weak in individual scans, but coherently adds as we add scans. Figure C.2 shows a plot similar to what was presented in the previous section, but that stacks all available Level 2 scans on the beginning of each scan. For Feed 15, we see a very strong wave across frequency, which falls to zero around 17 seconds after the start of the scan, and then switches from negative to positive, or positive to negative power. This is simply an artifact of the low-pass normalization we perform during TOD processing, and the strong wave at the very beginning of each scan is the real start-of-scan feature. This artifact explains why we, in the second and third rows of Figure 10 (where the start-of-scan effect can be seen), observe a similar switching of power as we move from the right edge of the map and towards the center.

Looking at Feed 11 in Figure C.2 a much weaker, but similar, start-of-scan feature can be seen. Generally, all feeds show very similar behavior to either Feed 11 or 15: all Feeds associated with DCM1-2 (Feeds 6, 14, 15, 16, and 17) have very similar behavior, and all remaining Feeds show only a weak start-of-scan effect, as appears in Feed 11. It is unclear why this clear divide exists, and how it relates to DCM1-2.

The exact origin of the start-of-scan effect is unknown, but a standing wave induced by mechanical vibration is also a strong

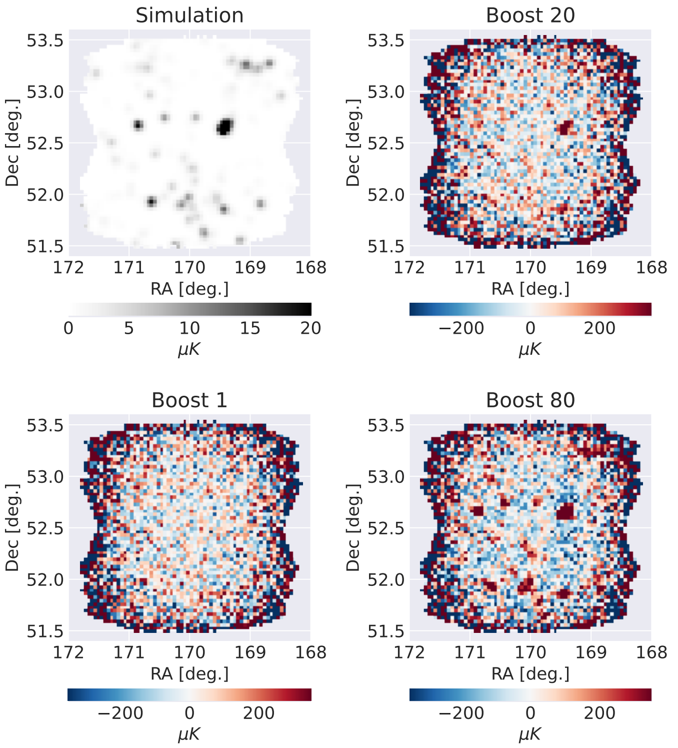


Fig. D.1. Illustration of the signal-injection method for transfer function analysis for different signal boost strengths. The top left plot shows the signal-only CO simulation over the relevant patch. The remaining plots show the resulting maps of the simulation injected into the real TOD with different boost strengths and passed through the entire COMAP pipeline. The boost is relative to the fiducial model of Chung et al. (2022), used in the simulations. All four plots show the same frequency slice centered at 26.953 GHz.

candidate for this systematic effect. The re-pointing that is performed in between scans is currently the only time in the scanning strategy the elevation drive is utilized, as our scans are performed in constant elevation mode.

As with the turn-around effect, some effort was made to model the start-of-scan effect in the time domain. This was fairly successful, and fitting a decaying exponential function to the beginning of each scan appeared to remove more than 90 % of the signal induced by this effect. However, the map-PCA proved much more effective than the time-domain efforts, and they are therefore not employed.

Appendix D: Signal injection example maps

Figure D.1 illustrates the signal injection pipeline. The first figure shows the CO simulation itself, while the subsequent three panels show the results of injecting this simulation into the real data with different boost strengths. Note that the simulation is injected into the TOD of the Level 1 data, and the maps shown have gone through the entire pipeline.

Appendix E: Uncertainty and frequency maps

Figure E.1 compares the uncertainties of the Season 2 maps to the Season 1 maps. The values are averages across all frequencies, calculated by inverse-variance co-addition of the uncertainties, as

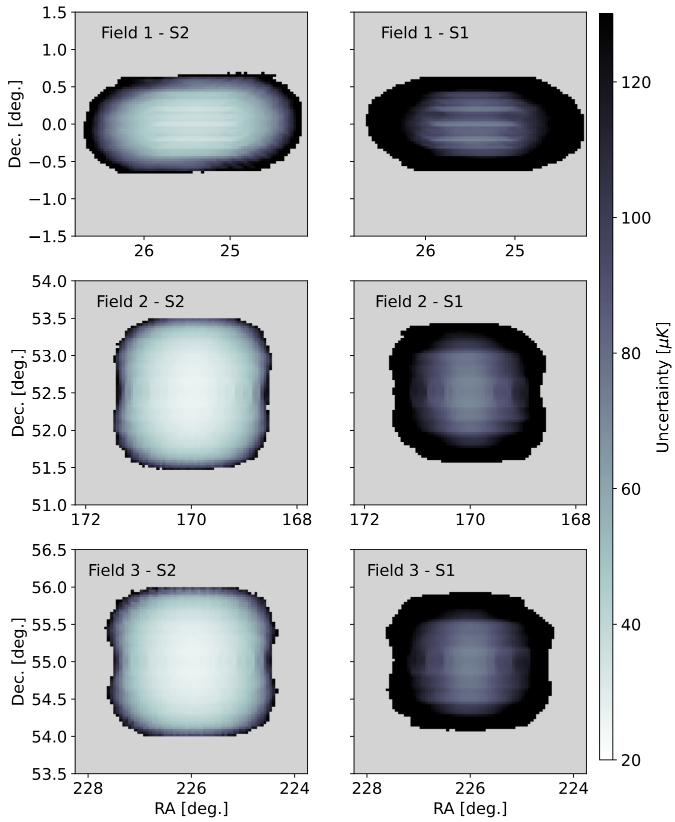


Fig. E.1. Uncertainties across the three fields for Season 2 (S2) maps published in this work (left), and the Season 1 (S1) maps published in ES (right).

$$\sigma^{\text{mean}} = \sqrt{\frac{1}{\langle 1/\sigma_\nu^2 \rangle}}. \tag{E.1}$$

All frequencies have relatively similar uncertainties, with some exceptions close to the Band edges. The center of the maps have a uncertainties of around 25 μK, while the high-sensitivity $\sim 1.5° \times 1.5°$ regions have an uncertainty of < 50 muK for Fields 2 and 3, with the Field 1 region being slightly smaller.

Figure E.2 shows feed-coadded individual frequency maps for Field 2 across 32 GHz – 34 GHz (1/4th of all channels) for all the data of Season 1 and 2 combined, processed with the Season 2 pipeline. All maps are noise dominated after the map-level PCA filtering. The noise increases towards the highest frequencies because of aliasing cuts on older data (see Sect. 2.3).

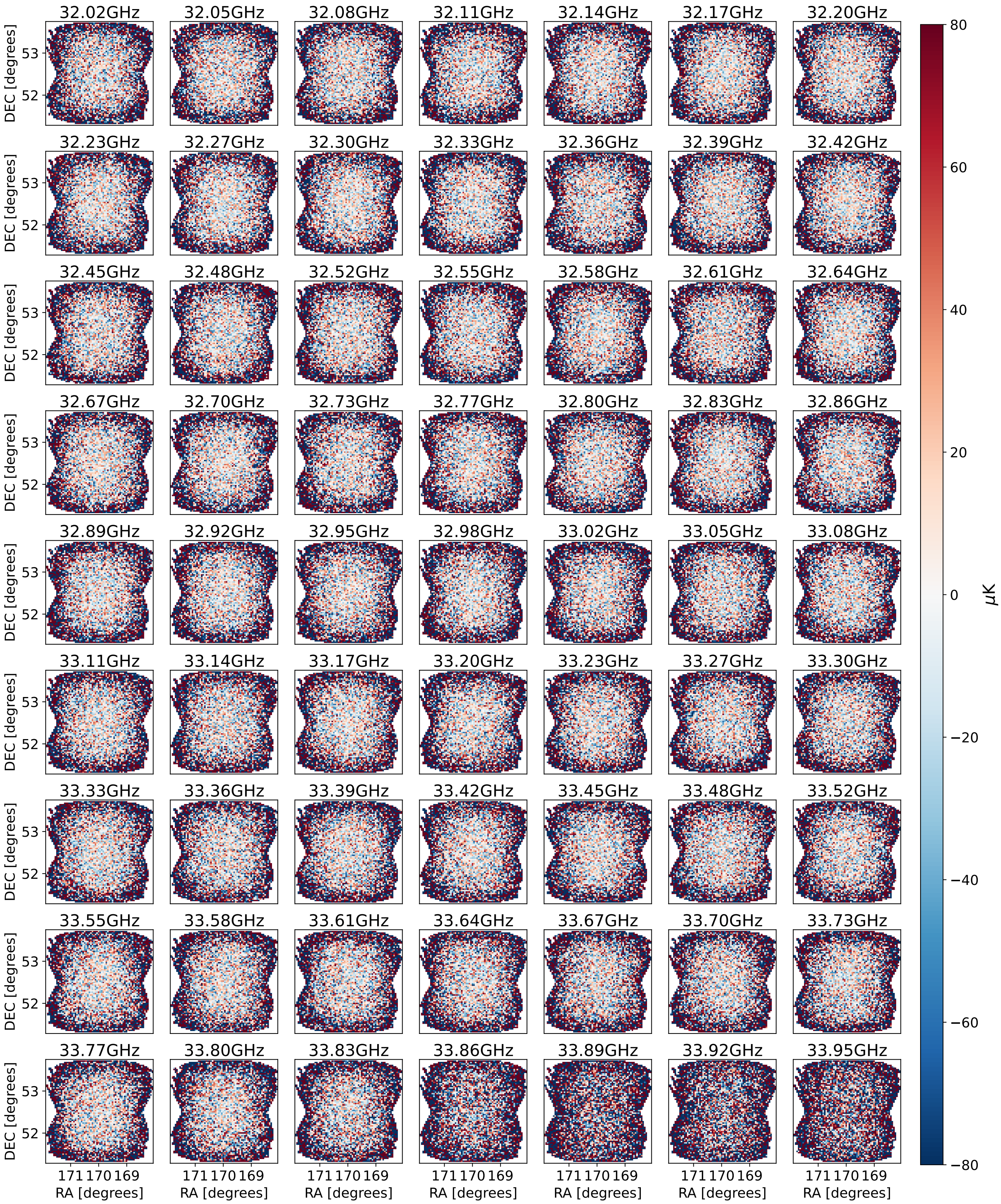


Fig. E.2. Final 31.25 MHz wide frequency maps for sideband A:USB and Field 2. The titles of each sub-plot indicate the center frequencies of each frequency map.

Paper IV

COMAP Pathfinder – Season 2 results II. Updated constraints on the CO(1–0) power spectrum

Astronomy & Astrophysics manuscript no. aanda
June 18, 2024

COMAP Pathfinder – Season 2 results
II. Updated constraints on the CO(1–0) power spectrum

N.-O. Stutzer[1,⋆], J. G. S. Lunde[1], P. C. Breysse[6,7], D. T. Chung[3,4,5], K. A. Cleary[2], D. A. Dunne[2], H. K. Eriksen[1], H. T. Ihle[1], H. Padmanabhan[8], D. Tolgay[3,14], I. K. Wehus[1], J. R. Bond[3,14,15], S. E. Church[17], T. Gaier[12], J. O. Gundersen[18], A. I. Harris[19], S. E. Harper[10], R. Hobbs[9], J. Kim[13], J. W. Lamb[9], C. R. Lawrence[12], N. Murray[3,14,15], T. J. Pearson[2], L. Philip[12,16], A. C. S. Readhead[2], T. J. Rennie[10,11], and D. P. Woody[9] (COMAP Collaboration)

[1] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, N-0315 Oslo, Norway
[2] California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
[3] Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada
[4] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[5] Department of Astronomy, Cornell University, Ithaca, NY 14853, USA
[6] Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY, 10003, U.S.A
[7] Department of Physics, Southern Methodist University, Dallas, TX 75275, USA
[8] Departement de Physique Théorique, Universite de Genève, 24 Quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
[9] Owens Valley Radio Observatory, California Institute of Technology, Big Pine, CA 93513, USA
[10] Jodrell Bank Centre for Astrophysics, Alan Turing Building, Department of Physics and Astronomy, School of Natural Sciences, The University of Manchester, Oxford Road, Manchester, M13 9PL, U.K.
[11] Department of Physics and Astronomy, University of British Columbia, Vancouver, BC Canada V6T 1Z1, Canada
[12] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109
[13] Department of Physics, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
[14] Department of Physics, University of Toronto, 60 St. George Street, Toronto, ON, M5S 1A7, Canada
[15] David A. Dunlap Department of Astronomy, University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada
[16] Brookhaven National Laboratory, Upton, NY 11973-5000
[17] Kavli Institute for Particle Astrophysics and Cosmology and Physics Department, Stanford University, Stanford, CA 94305, USA
[18] Department of Physics, University of Miami, 1320 Campo Sano Avenue, Coral Gables, FL 33146, USA
[19] Department of Astronomy, University of Maryland, College Park, MD 20742

Received MM DD, YYYY; accepted MM DD, YYYY

ABSTRACT

We present updated constraints on the cosmological 3D power spectrum of carbon monoxide CO(1–0) emission in the redshift range 2.4–3.4. The constraints are derived from the two first seasons of Carbon monOxide Mapping Array Project (COMAP) Pathfinder line-intensity mapping observations aiming to trace star-formation during the Epoch of Galaxy Assembly. These results improve on the previous Early Science (ES) results through both increased data volume and improved data processing methodology. On the methodological side, we now perform cross-correlations between groups of detectors ("feed-groups"), as opposed to cross-correlations between single feeds, and this new feed-group pseudo power spectrum (FGPXS) is constructed to be more robust against systematic effects. In terms of data volume, the effective mapping speed is significantly increased due to an improved observational strategy as well as better data selection methodology. The updated spherically- and field-averaged FGPXS, $\tilde{C}(k)$, is consistent with zero, at a probability-to-exceed of around 34 %, with an excess of 2.7 σ in the most sensitive bin. Our power spectrum estimate is about an order of magnitude more sensitive in our six deepest bins across $0.09\,\mathrm{Mpc}^{-1} < k < 0.73\,\mathrm{Mpc}^{-1}$, as compared to the feed-feed pseudo power spectrum (FPXS) of COMAP ES. Each of these bins individually constrains the CO power spectrum to $kP_{\mathrm{CO}}(k) < 2400 - 4900\,\mu\mathrm{K}^2\mathrm{Mpc}^2$ at 95 % confidence. To monitor potential contamination from residual systematic effects, we analyze a set of 312 difference-map null tests and find that these are consistent with the instrumental noise prediction. In sum, these results provide the strongest direct constraints on the cosmological 3D CO(1–0) power spectrum published to date.

Key words. galaxies: high-redshift – radio lines: galaxies – diffuse radiation – methods: data analysis – methods: observational

1. Introduction

By collecting the combined redshift-dependent line emission from all sources, both diffusely emitting gas and all galaxies, bright and faint, line intensity mapping (LIM) aims to map the Universe from large to small scales in three dimensions (see Madau et al. 1997; Battye et al. 2004; Peterson et al. 2006; Loeb & Wyithe 2008; Kovetz et al. 2017, 2019, and references therein for details on LIM). Several emission lines of interest have been proposed, among them 21 cm, carbon monoxide (CO), ionized carbon ([C II]), Lyα and Hα, each with different astrophysical

⋆ e-mail: n.o.stutzer@astro.uio.no

and cosmological goals (Kovetz et al. 2017, 2019; Bernal & Kovetz 2022).

At the forefront of CO LIM is the CO Mapping Array Project (COMAP), currently in its Pathfinder phase, which aims to measure the large-scale CO(1–0) line emission at redshifts $z \sim 2.4$–3.4, tracing the star-forming galaxies around the Epoch of Galaxy Assembly (Cleary et al. 2022). The COMAP Pathfinder instrument is a focal plane array of 19 detectors (which we refer to as "feeds") each with independent receiver electronics, fielded on a 10.4m Leighton telescope at the Owens Valley Radio Observatory. It observes in a frequency range of 26–34 GHz and is sensitive to 115.27 GHz CO(1–0) rotational line emission at redshift $z \sim 2.4$–3.4. Based on the first year of observations ("Season 1"), COMAP obtained the first direct limits on the 3D CO(1–0) clustering power spectrum, already ruling out several models from the literature. These results were published in a series of eight Early Science (ES) papers, along with a preview of our ongoing continuum survey of the Galaxy, a look at the prospects for CO LIM at the Epoch of Reionization, and a cross-correlation of ES data with an overlapping galaxy survey (Cleary et al. 2022; Lamb et al. 2022; Foss et al. 2022; Ihle et al. 2022; Chung et al. 2022; Rennie et al. 2022; Breysse et al. 2022; Dunne et al. 2024).

In this paper, the second of a series of three, we update our power spectrum results based on observations taken in our first and second seasons (S2), following Ihle et al. (2022). We build on the filtered and calibrated low-level COMAP data products described in detail by Lunde et al. (2024). Implications for astrophysical constraints and modeling are explored by Chung et al. (2024).

As discussed by Lunde et al. (2024), the current experimental design is overall very similar to ES, but takes into account a few important lessons learned. For example, COMAP Season 2 uses only constant elevation scans (CES), not Lissajous scans, because one of the main conclusions of Ihle et al. (2022) was that changes in elevation within a scan result in significant residual systematic effects from changes in the atmospheric and or ground pickup signals. We also avoid elevations that are strongly contaminated by ground radiation received in the sidelobes. In addition, the instrument drive speed was decreased around May 2022 in order to reduce the stress on the telescope (Lunde et al. 2024), and the effective instrumental properties therefore changed notably about halfway through the second season. We denote periods before and after the speed change the "fast-" and "slow-moving azimuth scans", respectively (and are equivalent to the naming convention "Season 1+Season 2a" and "Season 2b" used by Lunde et al. (2024), where "a" and "b" denote the period before and after the drive changes).

For consistency with previous COMAP publications, we adopt the same ΛCDM cosmological model as Chung et al. (2022) and Li et al. (2016) when converting distances in our map cubes from angular and spectral frequency units into physical units. Explicitly, we set $\Omega_m = 0.286$, $\Omega_\Lambda = 0.714$, $\Omega_b = 0.047$, $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ where $h = 0.7$, $\sigma_8 = 0.82$ and $n_s = 0.96$, which is roughly consistent with WMAP (Hinshaw et al. 2013). Unless otherwise stated all distances and distance-derived quantities in megaparsecs carry an implicit h^{-1}.

This paper is structured as follows: the power spectrum methodology and updated null test framework are presented in Sect. 2 and 3 respectively. In Sect. 4 we present the power spectrum transfer function used to account for signal loss from low-level filtering and instrumental effects. Sections 5 and 6 show the power spectrum results and the outcome of our null tests. Our conclusions are presented in Sect. 7.

Table 1. Feed-groups used in the feed-group pseudo cross-power spectrum.

DCM1 (feed-group)	Feeds
1	1, 4, 5, 12, 13
2	6, 14, 15, 16
3	2, 7, 18, 19
4	3, 8, 9, 10, 11

Notes. "Feed groups" and their associated first down-conversion (DCM1) electronics.

2. Power spectrum methodology

The power spectrum fully characterizes the information contained in a Gaussian random field and so is one of the most powerful statistics for cosmological density fields. While the non-linear physics of galactic emissions to which COMAP is sensitive is not fully Gaussian, the power spectrum is still a useful statistic, and complementary to other summary statistics such as the Voxel Intensity Distribution (VID); (Breysse et al. 2017; Ihle et al. 2019) or the Deconvolved Distribution Estimator (DDE); (Breysse et al. 2023; Chung et al. 2023).

The COMAP Pathfinder uses three-dimensional maps of the CO(1–0) emission to constrain models of star formation during the Epoch of Galaxy Assembly. While the maps already represent the compression of hundreds of terabytes of raw time-ordered data (TOD) into only a few gigabytes, it is possible to encode and compress much of the relevant astrophysical and cosmological information contained within the maps even more by using summary statistics like the power spectrum. As such the power spectra are easier and more computationally efficient to work with when constraining astrophysical and cosmological information of the mapped emission field.

In the COMAP ES paper series, Ihle et al. (2022) devised a novel cross-power spectrum methodology, the feed-feed pseudo cross-power spectrum (FPXS), constructed to be robust against systematic errors. This work largely builds on the methodology developed by Ihle et al. (2022) and lessons learned since the ES data processing to improve the power spectrum constraints of COMAP even further. In the following we summarize the FPXS methodology used and outline what has changed from the methodology developed by Ihle et al. (2022).

2.1. The Feed-Group Pseudo Cross-Power Spectrum

We begin by defining the general concepts of an auto- and cross-power spectrum. The auto-power spectrum can simply be defined as the variance of Fourier modes of a map. It can be written as

$$P(\boldsymbol{k}) = \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}}}\langle|\mathcal{F}\{\boldsymbol{m}_i\}|^2\rangle = \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}}}\langle|f_i(\boldsymbol{k})|^2\rangle, \quad (1)$$

where V_{vox} is the volume of a voxel (i.e. three-dimensional pixel) in units Mpc^3, N_{vox} is the number of voxels and $f_i(\boldsymbol{k})$ are the Fourier coefficients of the map $\boldsymbol{m}_i$ at wavenumber $\boldsymbol{k}$, in units Mpc^{-1}. For the Fourier transform $\mathcal{F}\{\boldsymbol{m}_i\}$ of the map $\boldsymbol{m}_i$ we use the same convention previously used in ES (Ihle et al. 2022; Harris et al. 2020). We can safely use the regular Fourier basis in the case of COMAP, instead of the more general spherical harmonics, as the fields are only $\sim 2°$ in diameter and the flat-sky approximation is sufficient. Note that, since our maps are three-dimensional, so is the power spectrum derived from those maps.

The auto-power spectrum will pick up all components that contribute to the variance in the map: signal, noise and systematic effects. It can thus be decomposed into

$$P(\boldsymbol{k}) = P_{\mathrm{CO}}(\boldsymbol{k}) + P_{\mathrm{noise}}(\boldsymbol{k}) + P_{\mathrm{syst}}(\boldsymbol{k}), \quad (2)$$

showing contributions from CO signal[1], noise and systematic effects, respectively. To obtain an unbiased estimate of the signal power spectrum, the systematic effects and noise properties of the map have to be understood and modeled.

Similarly to the auto-power spectrum, we can define a cross-power spectrum between two maps to be the covariance of Fourier coefficients of the two. It can be written as

$$C_{ij}(\boldsymbol{k}) = \frac{V_{\mathrm{vox}}}{N_{\mathrm{vox}}} \left\langle \mathrm{Re}\{f_i^*(\boldsymbol{k}) f_j(\boldsymbol{k})\} \right\rangle, \quad (3)$$

where f_i and f_j represent Fourier coefficients of two different maps i and j. The cross-power spectrum reduces to the auto-power spectrum if the two maps i and j are chosen to be identical.

As opposed to the auto-power spectrum, a cross-power spectrum will only be sensitive to correlated common modes between the two maps. Independent noise and independent systematic effects will therefore be canceled out and we can decompose the cross-spectrum as

$$C_{\mathrm{ij}}(\boldsymbol{k}) = P_{\mathrm{CO}}(\boldsymbol{k}) + C_{ij,\mathrm{common}}(\boldsymbol{k}), \quad (4)$$

where $C_{ij,\mathrm{common}}(\boldsymbol{k})$ represents the cross-spectrum contribution from some common systematic effect between the two maps.

We can now see a powerful property of the cross-spectrum: if we can choose two maps with independent noise properties and statistically unique systematic effects, giving $C_{ij,\mathrm{common}}(\boldsymbol{k}) = 0$, the cross-power spectrum will yield an unbiased estimator for the signal power spectrum $P_{\mathrm{CO}}(\boldsymbol{k})$.

This is the main property that the feed-feed pseudo cross-power spectrum (FPXS) developed by Ihle et al. (2022) is built to exploit. Because the COMAP Pathfinder measures the sky with 19 feeds, each with its own receiver signal chain, the maps from different detectors will have independent noise properties. Additionally, several systematic effects are believed to be unique to each feed, or specific group of feeds. Therefore, a cross-spectrum between two detector maps will not be biased by the noise contribution of the detectors or feed-specific systematic contamination.

In this work, rather than cross-correlating individual feeds, we instead cross-correlate groups of feeds. In particular, we group feeds by their shared first down-conversion (DCM1) local oscillator (Lamb et al. 2022). Table 1 shows the feeds that are grouped together in a given "feed-group".

The reason for this change is that some of the systematic effects uncovered with the improved sensitivity of the current data volume are correlated with the DCM1 feed-groups as shown by Lunde et al. (2024). Applying the original FPXS, involving cross-correlation of feeds from the same feed-group, would not have been effective in canceling such systematic effects since they are common-mode for a given feed-group.

Instead, by grouping all feeds in a given feed-group together, detectors from the same feed-group are never cross-correlated when computing the average feed-group pseudo cross-power spectra (FGPXS). This effectively cancels the systematic effects that are common to each feed-group, while retaining the CO signal. Additionally, grouping together detectors in this way produces effective detector maps that have more sky overlap. Thus when cross-correlating these maps we obtain better constraints on large-scale power spectrum modes and less mode-mixing due to a larger cross-map footprint. The result from a lower degree of mode mixing is is also a lower amount of large-scale systematic effects that can leak into the small- and intermediate-scale power. This is especially important, as we know from Lunde et al. (2024) that our most dominant systematic effects are large-scale modes in the maps.

However, even though the FGPXS is slightly more robust to systematic effects, this comes at the price of a slight decrease in sensitivity. The expected loss in sensitivity when using FGPXS as opposed to FPXS should in theory follow the upper limit found by Ihle et al. (2022):

$$\sigma_{C(\boldsymbol{k})}^{N_{\mathrm{split}}} \geq \sqrt{\frac{1}{1 - 1/N_{\mathrm{split}}}} \sigma_P(k), \quad (5)$$

where the uncertainty of a cross-spectrum, $\sigma_{C(\boldsymbol{k})}^{N_{\mathrm{split}}}$, is given by the number of cross-correlated data splits, N_{split}, compared to the optimal sensitivity, $\sigma_P(k)$, one can obtain when using all available data in an auto-power spectrum. To give some intuition on Eq. (5), we show a grid of possible feed-group and elevation split combinations. Equation (5) can be obtained from the ratio between the total number of split combinations (i.e the optimal auto-spectrum sensitivity σ_P) and the number of all cross-combinations that do not constitute auto-combinations between feeds or elevations (respectively dark and light gray shading). From this, we should expect there to be a loss in sensitivity in the FGPXS compared to FPXS of $\sim 12\,\%$[2].

Nevertheless, we conservatively cluster feeds into the aforementioned feed-groups to avoid systematic effect contamination, at the price of a minor loss in sensitivity. Note that, apart from the reasons stated above, there is in principle no difference between the FPXS and FGPXS algorithmically; for instance it would be trivial to group the feeds in a different configuration if that were found to be advantageous in the future for some reason. We can thus describe the two methods using the same algorithmic representation shown in the following.

After splitting the data into feeds or feed-groups, we split the data additionally into halves, each with independent systematic effect contributions, e.g. high/low elevation as done by Ihle et al. (2022), which further eliminates unwanted contributions to the cross-spectrum. We can write the main steps of the FGPXS algorithmically as follows.

1. Split the data into two halves A and B. As done by Ihle et al. (2022), we chose elevation as the main cross-correlation variable to eliminate potential sidelobe pickup from the ground.
2. For parts A and B respectively make maps of each feed-group i. We denote these by, e.g., m_{A_2} for a map of part A with feed-group 2.
3. For each combination of feed-groups i and j, and data splits A and B, compute cross-power spectra.

[1] Note that technically $P_{\mathrm{CO}}(\boldsymbol{k})$ in this notation would include contributions from both cosmic CO and all other astrophysical components with non-trivial frequency structure that are not subtracted out by the low-level data analysis steps, e.g. potential interloper line emission. However, for CO(1–0) at $z = 2$–3 there are very few, if any, interloper lines that could be picked up and we therefore use a CO-only notation.

[2] Note that this number in practice tends to be a little larger because we exclude auto-combinations between feed(-groups) and these contain the largest fraction of the optimal total auto-spectrum sensitivity due to better overlapping cross-sky maps.

4. Compute a noise weighted average FGPXS of all the resulting $N_{\text{feed-group}} \times (N_{\text{feed-group}} - 1)$ (with $N_{\text{feed-group}} = 4$ when using feed-groups and $N_{\text{feed-group}} = 19$ if computing the ES FPXS) individual FGPXS that do not involve the same detector or elevation:

$$C(\boldsymbol{k}) = \frac{\sum_{A_i \neq B_j} \frac{C_{A_i B_j}(\boldsymbol{k})}{\sigma^2_{C_{A_i B_j}}(\boldsymbol{k})}}{\sum_{A_i \neq B_j} \frac{1}{\sigma^2_{C_{A_i B_j}}(\boldsymbol{k})}}, \tag{6}$$

with corresponding uncertainty

$$\sigma_{C(\boldsymbol{k})} = \frac{1}{\sqrt{\sum_{A_i \neq B_j} \frac{1}{\sigma^2_{C_{A_i B_j}}(\boldsymbol{k})}}} \tag{7}$$

This is what we refer to as the mean feed-group pseudo cross-power spectrum (FGPXS) or feed-feed pseudo cross-power spectrum (FPXS), if (respectively) feed-groups or feeds are used as effective detectors.

In Fig. 1 we illustrate a grid of possible feed-group and elevation combinations used for an average FGPXS. Those shaded dark and light gray represent auto-feed and auto-elevation combinations (respectively). The combinations that cross neither feed-groups nor elevations, indicated with examples of 2D FGPXS combinations, are used in the final average FGPXS in Eq. (6).

Due to the non-uniform coverage of our sky fields, as well as a non-trivial survey footprint (see Lunde et al. 2024, for examples of maps), the maps are weighted prior to computing their Fourier coefficients. We use the same weighting scheme as Ihle et al. (2022), given for a cross-power spectrum by

$$\boldsymbol{w}_{A_i B_j} \propto \frac{1}{\boldsymbol{\sigma}_{A_i} \boldsymbol{\sigma}_{B_j}}, \tag{8}$$

where $\boldsymbol{\sigma}_{Y_x}$ represents the uncertainty estimate in each voxel of a feed-group and elevation split map $\boldsymbol{m}_{Y_x}$. These weights are then applied to the map, $\tilde{\boldsymbol{m}}_i = \boldsymbol{w}_i \boldsymbol{m}_i$, before power spectrum estimation with the Fourier coefficients $\tilde{f}_i(\boldsymbol{k}) = \mathcal{F}\{\boldsymbol{w}_i \boldsymbol{m}_i\}$ in Eq. (3). Regions outside the map footprints are assigned zero weights. The power spectra of these weighted maps are commonly referred-to as pseudo power spectra (Hivon et al. 2002). The pseudo power spectra are a biased power spectrum estimator because different Fourier modes become coupled via the applied weights (see Hivon et al. 2002; Leung et al. 2022, for details on mode mixing). Note that we will use $\tilde{P}(k)$ to denote pseudo spectra in the later results sections, but we use the notation $P(k)$ (without the tilde) in the methods sections as most of the methodology is equivalently written for unbiased and pseudo spectra. A detailed discussion of the COMAP-specific mode mixing can be found in Fig. 1 and Appendix D of Ihle et al. (2022), which shows that the effect is $\leq 20\,\%$ over our k-range. Reversing the mode-mixing will thus be left as a future exercise and is beyond the scope of this work.

2.2. The binned power spectrum estimator

As COMAP produces line-intensity maps spanning three-dimensional redshift-space volumes, the resulting power spectra also span three-dimensional Fourier-space volumes. It can, however, be easier to work with a power spectrum spherically averaged down to one dimension. For the spherically-averaged

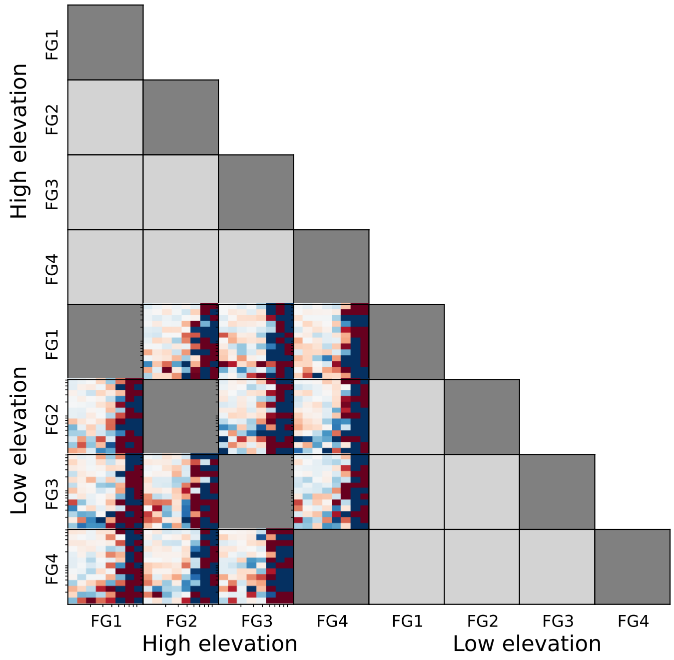


Fig. 1. Example grid of possible feed-group (FG1–4) and elevation (high and low) split combinations. Combinations with dark and light gray shading, respectively, represent auto-feed-group and auto-elevation combinations which are not used in the final averaged FGPXS. The cross-combinations containing examples of a 2D cross-spectrum are kept in the final average FGPXS as neither identical feed-groups, nor elevations, are crossed.

power spectrum to contain all relevant information in the full three-dimensional power spectrum, the emission field is technically required to remain statistically isotropic on large scales and stationary across the mapped redshift range. This is not strictly known to be true for the CO emission field. Cosmic star-formation, especially dust-obscured star-formation history traced in the IR, is poorly constrained in our targeted redshift range of $z = 2.4$–3.4 (see Madau & Dickinson 2014, for a review of cosmic star-formation history). As a consequence the extent to which the mapped CO emission field is stationary is largely unknown. The spherically-averaged power spectrum of a dynamic field will not be sensitive to changes of the CO emission across cosmic time, but it will measure the time-averaged properties of the targeted CO structures. However, the distinction is moot for the current COMAP signal-to-noise ratio (SNR), as no clear CO excess is observed in the power spectra. Thus we present the spherically-averaged, 1D power spectrum as our main science product.

Additionally, in practice, the angular and the redshift axes are observed in fundamentally different ways, and the low-level filtering applied to the data (Lunde et al. 2024) as well as redshift space distortions and line-broadening (Chung et al. 2021) can affect the signal and sensitivity differently along each axis. Therefore, the angular and line-of-sight dimensions are convenient to separate, and we bin the 3D power spectrum $C(\boldsymbol{k})$, with $\boldsymbol{k} = (k_x, k_y, k_z)$, into both a cylindrical and spherically-averaged power spectrum. The former of these conserves the structures perpendicular and parallel to the line-of-sight by only merging the two angular axes

$$\boldsymbol{k}_i = (k_\perp, k_\parallel) = \left(\sqrt{k_x^2 + k_y^2}, k_z\right). \tag{9}$$

Meanwhile, the latter will average the 3D power spectrum into 1D bins of the form

$$k_i = \sqrt{k_x^2 + k_y^2 + k_z^2}. \tag{10}$$

The binned cylindrically-averaged power spectrum estimator will then become

$$C(\boldsymbol{k}) \approx C_{\boldsymbol{k}_i} = \frac{V_{\text{vox}}}{N_{\text{vox}} N_{\text{modes}}} \sum_{\boldsymbol{k} \in \boldsymbol{k}_i} \langle \text{Re}\{f_1^*(\boldsymbol{k}) f_2(\boldsymbol{k})\} \rangle , \tag{11}$$

where the number of Fourier modes in bin $\boldsymbol{k}_i$ is given as N_{modes}. The equation is completely analogous when binning to the spherically-averaged power spectrum. We henceforth refer to the cylindrically and spherically-averaged power spectrum estimators as "2D" and "1D" due to the number of axes needed to display them, but note that they still represent averages of a 3D density field.

The bin edges are chosen to cover the scales to which COMAP is most sensitive and correspond to those used in the COMAP ES power spectra (Ihle et al. 2022), but due to our large increase in sensitivity and better understanding of the origin of correlations on large angular scales we conservatively excise all perpendicular scales $k_\perp \lesssim 0.1\,\text{Mpc}^{-1}$ for this publication. On these scales we are dominated by sub-optimal cross-map overlap, which results in poor constraining power of the large-scale structure as well as the possibility of large-scale residual systematic effect leakage through mode mixing into the small-scale power spectrum modes (see Lunde et al. 2024, for examples of our map-domain systematic effects). In the future, we aim to recover the large scales at $k_\perp \lesssim 0.1\,\text{Mpc}^{-1}$. Lastly, we also mask the bins corresponding to the highest $k_\perp$ and $k_\parallel$ used by Ihle et al. (2022) to prevent issues with aliasing near the Nyquist frequency of the two respective dimensions: $k_\perp^{\text{Nyquist}} \approx 1.22\,\text{Mpc}^{-1}$ and $k_\parallel^{\text{Nyquist}} \approx 0.74\,\text{Mpc}^{-1}$.

2.3. Uncertainty estimation from randomized null maps

In order to compute the mean FGPXS and its errors, as shown in Eqs. (6) and (7), we need the power spectrum uncertainties for each feed-group and elevation cross-combination, $\sigma^2_{C_{A_iB_j}}(\boldsymbol{k})$. This can be done via two basic approaches: simulations and data-driven methods. Here, we first detail some problems with a simulation-based approach used previously in COMAP ES (Ihle et al. 2022) and subsequently argue for why a data-driven approach was chosen in this work.

In COMAP ES, Ihle et al. (2022) chose a simple simulation approach where the power spectrum uncertainties were computed from an ensemble of simple white noise maps, $\boldsymbol{m}_{\text{noise},i} \sim \mathcal{N}(0, \boldsymbol{\sigma})$, drawn from a zero-mean Gaussian distribution $\mathcal{N}$ with the voxel uncertainties $\boldsymbol{\sigma}$. These were then propagated to the power spectrum level. The main advantage of this approach is its computational efficiency. However, it can only reflect the white noise level within the map, while residual correlated noise and the effect of the pipeline filters on the noise will not be contained in the uncertainty from these simple white noise maps (see, for instance, the power spectral density (PSD) of TOD in Fig. 9 of Lunde et al. 2024, for an illustration of the noise properties of the filtered data). The simplified simulations proved an adequate method given the sensitivity of our ES data. With the increased sensitivity achieved at the end of S2, obtaining suitable power spectrum errors, σ_{C_k}, through simulations would require the sampling of noise from the time-ordered data (TOD) domain (ideally with additional ground-up modeling of all contributing systematic effects), propagating it all the way through the low-level pipeline (Lunde et al. 2024) up to the power spectrum. However, this would be computationally expensive because the low-level pipeline filters would have to be re-run for each ensemble, and require significant additional data modeling.

Given the drawbacks with both the white noise and a potential TOD-level simulation-based approach, a data-driven method was instead chosen for this work as it represents a relatively computationally inexpensive method of estimating the power spectrum uncertainties that automatically reflects all the properties of the data. In particular, we will draw from the simple idea that we can cancel the signal and systematic effects in a subtraction between data-half maps, while leaving the correct noise properties. In our case, we estimate σ_{C_k} by what we will refer to as an ensemble of Randomized Null Difference (RND) maps.

The first step in the RND calculation is to divide the set of all scans in the data into two randomized halves, A and B, from which we subsequently make maps $\boldsymbol{m}_{A,i}^{\text{RND}}$ and $\boldsymbol{m}_{B,i}^{\text{RND}}$. This is done for all random split realizations i. Both $\boldsymbol{m}_{A,i}^{\text{RND}}$ and $\boldsymbol{m}_{B,i}^{\text{RND}}$ should contain the same signal, and due to the randomization of the splits also the same systematic effects. Hence we can cancel both the signal and systematic effects by computing the difference between the two maps;

$$\Delta\boldsymbol{m}_i^{\text{RND}} = \frac{\boldsymbol{m}_{A,i}^{\text{RND}} - \boldsymbol{m}_{B,i}^{\text{RND}}}{2}. \tag{12}$$

The difference maps $\Delta\boldsymbol{m}_i^{\text{RND}}$ now optimally capture the white and correlated noise properties and biases (from low-level filters, the instrumental beam, etc.) of the real maps, but are without any of the signal or systematic effects. As such they reflect the true properties of the data to a high degree.

Finally, to obtain the uncertainty of the power spectrum σ_{C_k} we need to compute the FGPXS of each difference map $\Delta\boldsymbol{m}_i^{\text{RND}}$. From the resulting ensemble of such feed-group cross-spectra, $C_{\Delta m_i}^{\text{RND}}(\boldsymbol{k})$, we can compute the uncertainties $\sigma_{\text{RND}}(\boldsymbol{k})$ by taking the standard deviation over the ensemble. These can then be used when co-adding together feed-group spectra to obtain the final mean FGPXS as explained in Eqs. (6) and (7).

3. Power spectrum null tests

With the increased effective COMAP data volume and the resulting increased sensitivity comes the need for more effective null tests to ensure the data quality of our final power spectra.

As we explain in this section, the null tests devised in this work are based on difference maps in a similar way to the RND method used for uncertainty estimation described earlier in Sect. 2.3, except we are now splitting the maps on meaningful parameters instead of randomly. The goal then becomes finding null variables (e.g. high/low humidity or left/right moving scans; see Table C.2 for list of all chosen variables) that correlate to systematic effects in one of the null variable halves by which we split the data.

We can write the difference map of some null variable j as

$$\Delta\boldsymbol{m}_j^{\text{null}} = \frac{\boldsymbol{m}_{A,j} - \boldsymbol{m}_{B,j}}{2}, \tag{13}$$

where the maps $\boldsymbol{m}_{A,j}$ and $\boldsymbol{m}_{B,j}$ represent the maps of the two halves of the data respectively. If the chosen null variable correlates to a systematic effect, the difference map $\Delta\boldsymbol{m}_j^{\text{null}}$ will contain the systematic effect but cancel the signal. The difference maps can then be used to perform a null test, with the null hypothesis being that the null maps are consistent with the general noise properties of the maps. The associated voxel uncertainty of the null map is then given by

$$\sigma^{\text{null}}_{\Delta \boldsymbol{m}_j} = \frac{\sqrt{\sigma^2_{m_{A,j}} + \sigma^2_{m_{B,j}}}}{2}, \tag{14}$$

for uncertainties $\boldsymbol{\sigma}_{m_{A,j}}$ and $\boldsymbol{\sigma}_{m_{B,j}}$ of the maps $\boldsymbol{m}_{A,j}$ and $\boldsymbol{m}_{B,j}$ respectively.

For each of the null variables j we then take the difference between the two maps as described by Eqs. (13) and (14). As we use a cross-elevation FGPXS we must compute a difference map for both high and low elevation. The data are therefore split into four parts, two elevation ranges and two null variables halves, where we subtract across the latter in the map domain and cross-correlate the resulting null maps across the former using the FGPXS method described earlier. With the set of resulting null test FGPXS $C^{\boldsymbol{k}_i}_{\Delta m_j}$ we can perform a χ^2-test, with a null hypothesis that the difference maps are consistent with noise, by first computing

$$\chi^2_{\text{null},j} = \sum_{\boldsymbol{k}_i} \left(\frac{C^{\boldsymbol{k}_i}_{\Delta m_j} - \mu^{\boldsymbol{k}_i}_{\Delta m_j}}{\sigma_{C^{k_i}_{\Delta m_j}}} \right)^2 = \sum_{\boldsymbol{k}_i} \left(\frac{C^{\boldsymbol{k}_i}_{\Delta m_j}}{\sigma_{C^{k_i}_{\Delta m_j}}} \right)^2, \tag{15}$$

with the expectation value of the null FGPXS $\mu^{\boldsymbol{k}_i}_{\Delta m_j} = 0$ under the null hypothesis. Here $\sigma_{C^{k_i}_{\Delta m_j}}$ is the uncertainty of the null FGPXS $C^{\boldsymbol{k}_i}_{\Delta m_j}$ in bin k_i for null variable j, which is estimated using the RND method described earlier in Sect. 2.3.

Thereafter we can compute the probability-to-exceed (PTE) which quantifies the probability to obtain a value $\chi^2_{\text{null},j}$ or higher. The PTE is defined as

$$\text{PTE}(\chi^2) = 1 - \text{CDF}(\chi^2), \tag{16}$$

where for a given probability distribution function $P(\chi^2)$ of the $\chi^2_{\text{null},j}$ values the corresponding cumulative distribution function is denoted as $\text{CDF}(\chi^2)$.

In our case, $P(\chi^2)$ does not follow the usual analytical χ^2-distribution because the noise properties of the FGPXS are not completely known analytically (see Watts et al. 2020; Nadarajah & Pogány 2016; Gaunt 2019, for some examples of how cross-spectrum noise properties can look). We thus compute the PTEs numerically by using an ensemble of RND maps equivalent to those we already use for estimating uncertainties as these will perfectly reflect the noise properties and biases in the data, as well as obey the null hypothesis. For each data processing run we compute 244 RND maps of which we use 61 for power spectrum uncertainty estimation and the remaining 183 for measuring the numerical χ^2-distribution.

4. Transfer functions

As described by Foss et al. (2022) and Ihle et al. (2022) the COMAP maps are not unbiased as the low-level filtering of the data, the binning of the data into voxels, and the finite resolution of the telescope beam will attenuate the signal in the maps. In this section, we explain how we de-bias our power spectrum estimates using transfer functions for each of the main effects that result in signal loss. The beam and voxel window smoothing of the signal is corrected using analytically computed transfer functions, while the low-level filtering attenuation is quantified using simulations. Details on how each transfer function is estimated are discussed by Lunde et al. (2024).

4.1. Power spectrum transfer functions

When performing a power spectrum analysis of our maps as described in Sect. 2 we obtain an estimate of the signal that is biased by several different effects. To see how the signal is biased we can write the FGPXS signal estimator as

$$C_{\boldsymbol{k}} = T(\boldsymbol{k})P^{\text{CO}}_{\boldsymbol{k}} = T_f(\boldsymbol{k})T_{\text{b}}(k_\perp)T_p(k_\perp)T_\nu(k_\parallel)P^{\text{CO}}_{\boldsymbol{k}}, \tag{17}$$

where the transfer function $T(\boldsymbol{k})$ is the product of the filter transfer function $T_f(\boldsymbol{k})$, the beam smoothing transfer function $T_{\text{b}}(k_\perp)$ as well as the pixel and spectral channel windows, $T_p(k_\perp)$ and $T_\nu(k_\parallel)$. The transfer function can be written in this multiplicative form in the Fourier domain because the low-level filtering and the smoothing of small-scale structures due to the instrumental beam and voxel window of the map grid can all (approximately) be expressed as a convolution in map domain. In Fig. 2 we show the full transfer function product $T(\boldsymbol{k})$, while the individual transfer function elements are shown in detail in Sect. 6. of Lunde et al. (2024).

Using our transfer function estimate we can de-bias the FGPXS by deconvolution;

$$P^{\text{CO}}_{\boldsymbol{k}} = \frac{C_{\boldsymbol{k}}}{T(\boldsymbol{k})}, \tag{18}$$

with the uncertainties of the signal estimator being affected in a similar manner,

$$\sigma^{\text{CO}}_{P_k} = \frac{\sigma_{C_k}}{T(\boldsymbol{k})}, \tag{19}$$

becoming large whenever the transfer function $T(\boldsymbol{k})$ becomes small.

5. Power spectrum results

In this section, we present the main power spectrum results of this paper. The raw data going into the power spectra are filtered, calibrated and binned into maps after a set of data selection steps which remove scans that are likely contaminated by systematic effects. This is described in detail by Lunde et al. (2024).

As one of the main lessons learned from COMAP ES was to employ only constant elevation scans (CES), and no longer use a Lissajous scanning strategy, the data presented here consist only of CES data (Foss et al. 2022; Ihle et al. 2022). Specifically, we include all data obtained up to November 2023, both the ES (Season 1) CES data as well as all data gathered in S2. The data volume obtained in S2 is, as explained by Lunde et al. (2024), effectively around eight times larger than the Season 1 CES data after data selection. In addition, the ES analysis of Season 1 data excluded several detectors that were either offline or excluded due to clear signs of systematic excess in reduced χ^2-tests or in visual inspections of feed cross-spectra. We are able to include these in the S2 analysis because all feeds were functioning during S2 and the map-domain PCA described in Lunde et al. (2024) strongly suppresses detector-specific systematic effects.

We note that in the ES analysis Ihle et al. (2022) removed feed-feed cross-spectra both through a reduced χ^2-test, and manual inspection of misbehaving feed combinations. Due to better

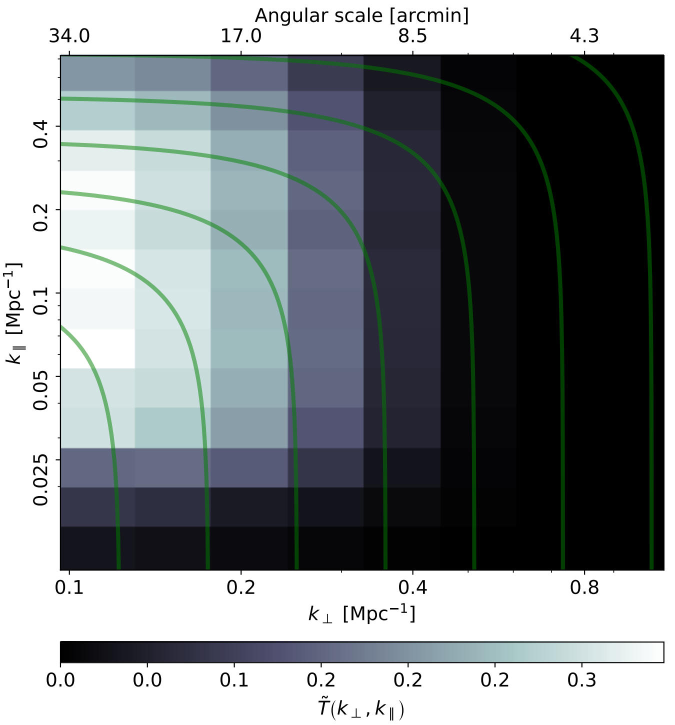


Fig. 2. Full power spectrum transfer function used to account for the signal losses due to the low-level filtering pipeline, the instrumental beam smoothing and the voxel window of the maps (see Lunde et al. 2024, for details on each individual transfer function). Thin green contours indicate the bin edges of the (1D) spherically-averaged FGPXS.

low-level data processing we were able to remove all data-driven cuts in the power spectrum domain, with the increased set of null tests since ES working as an additional safeguard against systematic effects (see Sect. (6) for discussion of null test results).

Lastly, in Appendix D, we show a simple end-to-end signal injection test as a qualitative test of our pipeline's ability to recover a known signal's amplitude within the estimated experimental errors and power spectrum transfer function.

5.1. The cylindrically averaged power spectrum result

In Fig. 3 we show the cylindrically-averaged (2D) mean FGPXS for all three fields separately, as well as in combination. The figure also shows the sensitivity per $(k_{\perp}, k_{\parallel})$-bin as well as the FGPXS in units of the sensitivity.

When looking at the 2D FGPXS in Fig. 3 we note that the noise blows up on small scales, particularly so in the angular direction, due to the COMAP transfer function seen in Fig. 2 (Lunde et al. 2024). However, we see no obvious patterns in the 2D FGPXS that would indicate a systematic effect. In fact, the spectra resemble white noise.

As mentioned earlier, in Sect. 2.2, we want to avoid issues with poorly constrained large-scale modes, strong mode mixing and possible residual large-scale systematic effects. We mitigate these issues by excluding 2D bins at $k_{\perp} < 0.1\,\mathrm{Mpc}^{-1}$. An example of spurious fluctuations induced by poor overlap can be seen in the COMAP ES cylindrically-averaged FPXS of Field 1 (see Ihle et al. 2022, noting that Field 1 is especially susceptible to poor detector overlap due to its position at declination zero) as correlated structures along constant $k_{\perp}$ at scales $k_{\perp} < 0.1\,\mathrm{Mpc}^{-1}$. These correlations have since been understood to originate from sub-optimal detector overlap, and are pushed to larger scales due to a larger sky overlap when computing cross-spectra between feed-groups instead of individual feeds. In interim estimates we found the average of the maximum correlations between a bin and all the others to be around 15 % at scales $k_{\perp} \geq 0.1\,\mathrm{Mpc}^{-1}$, while the correlations at $k_{\perp} \leq 0.1\,\mathrm{Mpc}^{-1}$ are somewhere in the 30–70 % regime. Improved modeling of these correlations will be the aim of future work.

5.2. The spherically-averaged power spectrum result

As interpreting the 2D cylindrically-averaged FGPXS can be somewhat unintuitive we can bin the spectra into 1D by performing a full spherical averaging. This is done as described in Sect. 2.2 where the 1D bin-edges are indicated as thin green contours in Fig. 3. When doing so we obtain the spherically-averaged FGPXS for the three fields, as well as the combination thereof, as seen in Fig. 4.

As discussed in Sect. 5.1, we excluded scales $k_{\perp} < 0.1\,\mathrm{Mpc}^{-1}$ from the power spectrum analysis to avoid issues with poor cross-map overlap, mode mixing and large correlations between large scale bins. Therefore, Fig. 4 only shows FGPXS data points on scales $k > 0.1\,\mathrm{Mpc}^{-1}$. Similar to the discussion in Sect. 5.1, we estimate the average of the maximum correlation between a 1D bin and all the others, on scales $k > 0.1\,\mathrm{Mpc}^{-1}$, to be $\lesssim 10\,\%$ after excluding the large $k_{\perp}$ scales and performing the spherical averaging. Given this $\lesssim 10\,\%$ level we will assume for Season 2 analyses downstream that the spherically-averaged 1D FPGXS bins are approximately uncorrelated. As with the 2D FGPXS discussed in Sect. 5.1, we intend to improve the exact modeling of these correlations in future work.

When looking at Fig. 4 we note that Fields 2 and 3 have the highest sensitivity, while Field 1 has around 50 % larger errors than the two other fields. This is because, of the three COMAP fields, Field 1 is most affected by the low-level data cuts (see Lunde et al. 2024), and due to its location at zero declination, is particularly susceptible to poor detector overlap. In addition, by rejecting poorly overlapping detector combinations, which are also the least sensitive, we prevent the resulting strong mode mixing and potential consequent leakage of systematic effects into smaller scales at the cost of a relatively minimal loss in sensitivity. As a consequence, we keep 2/3 of the feed-group cross-spectra of Field 1 in the analysis.

As we can see from Fig. 4, the cross-spectra are largely consistent with zero to within 2σ in most bins. However, perhaps the most notable feature is the high power in the second and most sensitive k-bin, at $0.12\,\mathrm{Mpc}^{-1} < k < 0.18\,\mathrm{Mpc}^{-1}$, which is respectively at around 2.3σ and 3σ significance above zero for Fields 1 and 2. Meanwhile for Field 3 the same bin is consistent with zero power. When combining the three fields, the co-added data point in the second k-bin has a value that is 2.7σ away from zero. For each of the spherically-averaged FGPXS of the three fields, we compute their χ^2 probabilities-to-exceed (PTE) to check their constancy with zero power. In doing so, we obtain PTEs of, respectively, 33.2 %, 19.5 % and 82.7 % for Fields 1–3. The field-combined spherically-averaged FGPXS results in a 34 % probability-to-exceed. As for the null tests, the PTE is estimated from the numerical RND χ^2 ensemble. While the combined 1D $\sim 2.7\sigma$ power in $0.12\,\mathrm{Mpc}^{-1} < k < 0.18\,\mathrm{Mpc}^{-1}$ bin is interesting, we do not consider it a statistically significant excess given the estimated PTEs. Thus we will have to wait for future analyses, and more data, to answer definitively whether this excess is simply noise or not.

Although we do not consider the field-combined 2.7σ point statistically significant, the agreement between two of the three

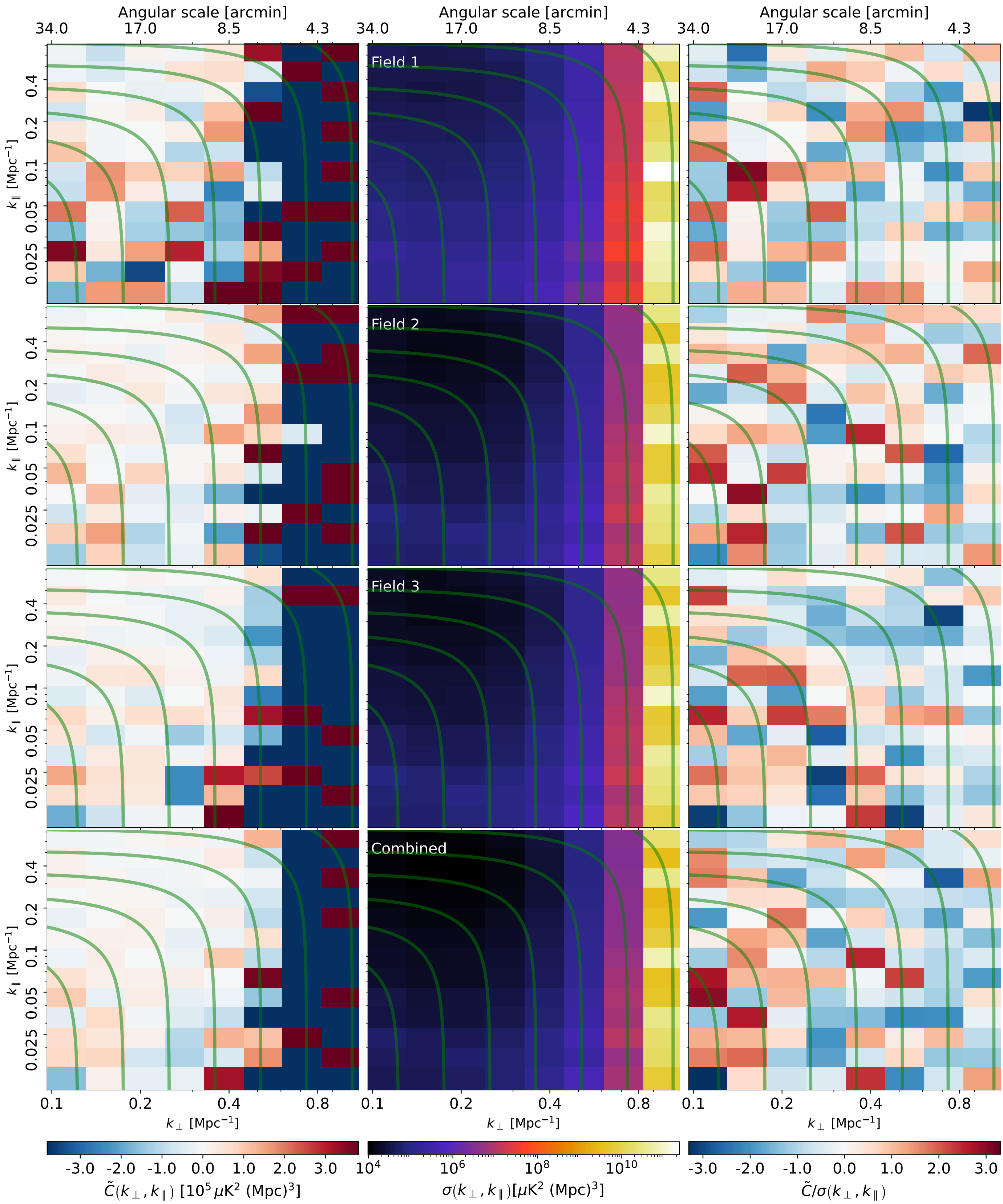


Fig. 3. Cylindrically-averaged (2D) feed-group pseudo cross-spectra. Columns show, from left to right, the full spectra, the corresponding 1σ uncertainty, and the ratio between the two. Rows show, from top to bottom, Fields 1, 2, 3, and all three combined. The approximate angular scale, assuming the central COMAP redshift at $z = 2.9$, corresponding to each $k_\perp$ is shown as a twin-axis on the upper row of plots. Thin green contours indicate the bin edges of the (1D) spherically-averaged FGPXS.

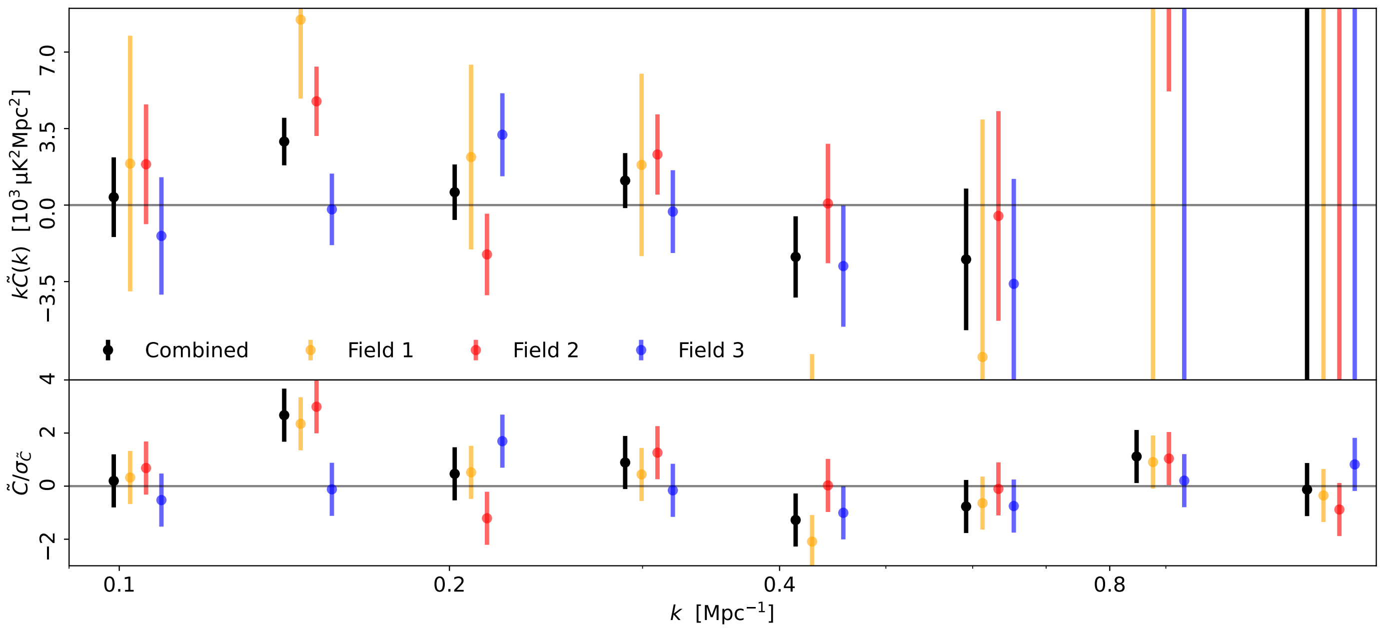


Fig. 4. Spherically-averaged FGPXS with 1σ uncertainties for Fields 1, 2 and 3 (orange, red and blue respectively) as well as the combination of the three (black) in units μK^{2}Mpc^2 (*upper panel*) and in units of the 1σ power spectrum uncertainties (*lower panel*). The data points have been slightly offset from their true k-position to increase readability (see Table B.1 in Appendix B for overview over bin centers, FGPXS values and uncertainties).

fields is interesting to note. As Fields 1 and 2 are quite different in terms of their path across the sky as seen from the telescope (Field 1 being a zero declination field while Field 2 is at a declination of 52.30°; see Foss et al. (2022) for details on the fields) they also are expected to have some independent systematic effects. However, Fields 2 and 3 are more alike and would be expected to share certain systematic effects.

5.3. Comparison to COMAP Early Science and COPSS

Having computed the spherically- and field-averaged FGPXS from the data we can compare it to the previous COMAP release as well as COPSS, the only other comparable CO(1–0) LIM survey in the literature with published data (Keating et al. 2016; Kovetz et al. 2017, 2019; Bernal & Kovetz 2022). This is illustrated in Fig. 5 where the field-averaged FGPXS is plotted together with the COMAP ES constant-elevation-scan FPXS of Ihle et al. (2022) and the individual COPSS data-points from Keating et al. (2016).

The first thing we notice when considering Fig. 5 is the dramatic reduction in the uncertainty of the current measurement compared to that from our ES phase (Ihle et al. 2022). Compared to the Ihle et al. (2022) FPXS the current level of sensitivity has increased by a factor $\sim$ 6–8 across our six most sensitive bins at $0.09\,\mathrm{Mpc}^{-1} < k < 0.73\,\mathrm{Mpc}^{-1}$. This illustrates the significant increase in the effective data volume by around a factor of eight. Even though the low-level data selection procedure detailed by Lunde et al. (2024) is somewhat more strict than the one in COMAP ES (Foss et al. 2022; Ihle et al. 2022), this is more than compensated by the lack of data cuts in the power spectrum domain, resulting in a significant increase in sensitivity overall. In other words, we have demonstrated that uncertainties in the power spectra integrate down in accordance with expectations for noise-dominated data.

The two highest k-bins have somewhat larger errors in the current result compared to the COMAP ES spectrum. This is due to a combination of the analytical beam transfer function now applied and a stricter 2D $\boldsymbol{k}$-space mask. The beam transfer function now applied is somewhat more strict than the numerically computed one of Ihle et al. (2022) on scales closer to the Nyquist limit in the angular direction. Additionally, to avoid problems with aliasing we have masked the outer-most bins in both $k_\parallel$ and $k_\perp$. As a result, the outer-most 1D k-bins contain a lower number of samples than they would have for the same 1D bins of Ihle et al. (2022).

The COPSS power spectrum estimate (Keating et al. 2016) primarily covers scales smaller than COMAP, but the two experiments overlap at $0.3\,\mathrm{Mpc}^{-1} \lesssim k \lesssim 1.0\,\mathrm{Mpc}^{-1}$, where they are largely consistent with each other. The only noteworthy disagreement between COPSS and the field-combined FGPXS is a mild $\sim 2.5\sigma$ tension in terms of the combined error between the two power spectrum estimates at $k \sim 0.6\,\mathrm{Mpc}^{-1}$. As we can see from Fig. 5 this point of mild tension coincides with one of the two COPSS points in which they reported a 2.5σ excess above zero.

Albeit with large uncertainties, we see that compared to COPSS and COMAP ES the updated COMAP data points cluster significantly closer to, and are consistent with, the two brightest models that were not already excluded in ES (Chung et al. 2022), i.e. the COMAP fiducial model[3] and the Li-Keating model of Keating et al. (2020). For more discussion of the consistency of COPSS with the current COMAP result, including modeling implications, we refer the interested reader to Chung et al. (2024).

When comparing the power spectrum sensitivity of COMAP to that of COPSS, we must take into account the smaller k-bins in the COPSS analysis. Although the two experiments have a certain region of overlap in k-space, the different bin sizes of

[3] A double power-law model relating halo masses in cosmological simulations to CO luminosities; see specifically “UM+COLDz+COPSS” in Table 1 of Chung et al. (2022) for their fiducial model definition.

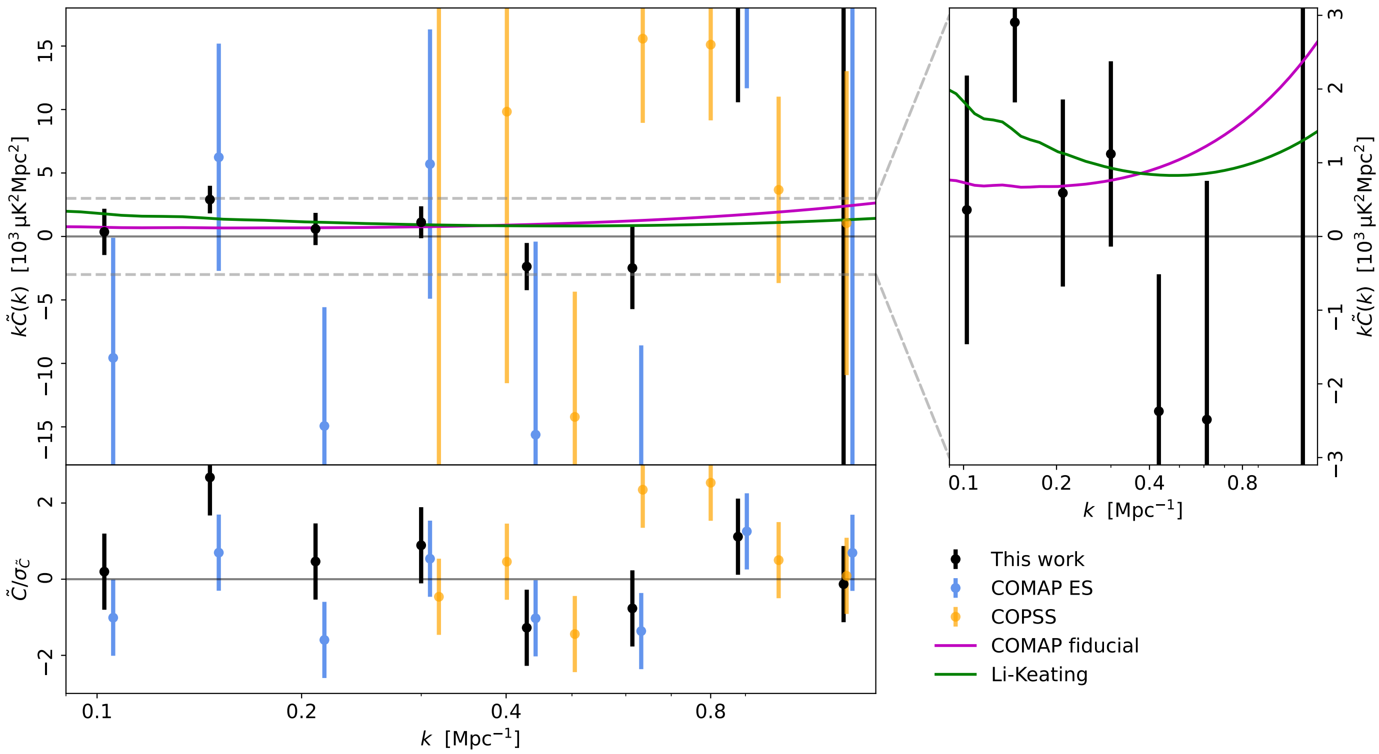


Fig. 5. (*Upper panel:*) Spherically-averaged FGPXS with 1σ uncertainties for the field-combined data presented in this paper (black), the COMAP ES field-averaged FPXS (blue; Ihle et al. 2022), and the COPSS power spectrum (orange; Keating et al. 2016). (*Lower panel:*) Corresponding power spectra divided by their respective $1\,\sigma$ uncertainty. (*Inset:*) Zoom-in of the COMAP data points and two comparable models from the literature, namely the fiducial second season COMAP model (Chung et al. 2022) and the Li-Keating model Li et al. (2016); Keating et al. (2020) model. None of the models includes any line-broadening discussed by Chung et al. (2021). Our data points and those of COMAP ES have been slightly offset from their true k-position to increase readability (see Table B.1 in Appendix B for overview over bin centers, FGPXS values and uncertainties).

COPSS and COMAP result in a different intrinsic within-bin variance. To mitigate this effect we can define the normalized sensitivity $\xi_k = \sigma_k \sqrt{\Delta V_k}$, where ΔV_k is the volume of a spherical k-shell defined by the bin k in k-space. Two bins with the same value for ξ_k would have the same sensitivity, σ_k, if they were binned to a standardized bin size. In other words, ξ_k traces the underlying continuous sensitivity of each experiment and k-scale, and as a result, the normalized sensitivity across k-bins and surveys becomes comparable. We illustrate the normalized sensitivity, ξ_k, in Fig. 6 for all data points shown in Fig. 5. The figure nicely illustrates the scales to which COMAP and COPSS are most sensitive. We see that COMAP is most sensitive on large scales at $0.1\,\mathrm{Mpc}^{-1} < k < 0.3\,\mathrm{Mpc}^{-1}$, while COPSS is most sensitive on small scales, $0.5\,\mathrm{Mpc}^{-1} < k < 1.0\,\mathrm{Mpc}^{-1}$, where the COMAP beam starts to dominate. However, the current FGPXS result has a peak sensitivity increase of around a factor of eight and ten compared to COMAP ES and COPSS respectively (see Table A.1 in Appendix A for a detailed list of exact normalized sensitivity improvements). This improvement in relative sensitivity compared to COPSS and COMAP ES is expected to increase further as the COMAP instrument gathers more data, and illustrates our ability to remove systematic effects to below the noise level and integrate the noise of the incoming data. In fact, as COPSS to-date remains the only comparable CO(1–0) LIM experiment with published data, COMAP currently provides the most sensitive CO(1–0) LIM constraints in the field.

5.4. Upper limits on the power spectrum

Given the factors of, respectively, eight and ten times the sensitivity of our power spectrum result compared to the COMAP ES and COPSS data, it is interesting to consider the upper limits (UL) at 95 % confidence on a non-zero CO(1–0) power spectrum that can be derived from the data-points. These are shown in Fig. 7 for the spherically- and field-averaged COMAP FGPXS, the COMAP ES (Ihle et al. 2022) data-points as well as the COPSS (Keating et al. 2016) power spectrum estimate. As we are at the level of sensitivity where it becomes more informative to look at the ULs per k-bin we only consider the ULs derived per k-bin in this work. We show only bin-wise derived ULs from the COMAP ES (Ihle et al. 2022) and COPSS (Keating et al. 2016) data to facilitate a direct comparison to our result. All ULs are computed under the assumption that the astrophysical CO signal must be positive. For comparison, two of the closest models from the literature are included in the plot; the COMAP fiducial model from Chung et al. (2022) and the Li-Keating model – a version of the Li et al. (2016) model from Keating et al. (2020). Note that, while the $0.12\,\mathrm{Mpc}^{-1} < k < 0.18\,\mathrm{Mpc}^{-1}$ FGPXS bin has a 2.7σ excess above zero, we still present it as a 95 % upper limit in Fig. 7 as we do not consider the excess statistically significant.

As in Fig. 6, the ULs we present in Fig. 7 reflect k-regions in which each survey is most sensitive. The 95 % ULs of this work and those derived from COMAP ES (Ihle et al. 2022) are most constraining in the six most sensitive COMAP bins at $0.09\,\mathrm{Mpc}^{-1} < k < 0.73\,\mathrm{Mpc}^{-1}$. Meanwhile

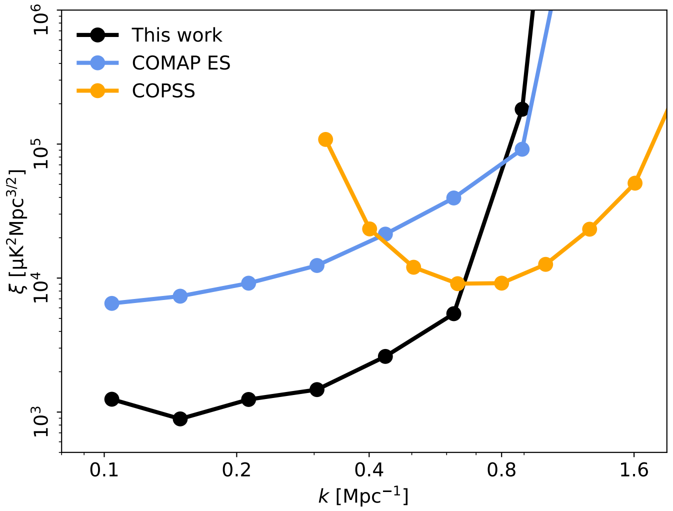


Fig. 6. Comparison of volume-normalized sensitivity, ξ_k, for the new COMAP FGPXS (black); the previous COMAP ES FPXS (blue; Ihle et al. 2022), and COPSS (orange; Keating et al. 2016). The deepest COMAP k-bin is roughly an order of magnitude more sensitive than the deepest COMAP ES and COPSS bins; for tabulated values, see Table A.1.

the COPSS Keating et al. (2016) ULs are at their lowest around $0.7\,\mathrm{Mpc}^{-1} < k < 1.6\,\mathrm{Mpc}^{-1}$, beyond where the COMAP beam and voxel window dominate and blow up the noise. Compared to COMAP ES, we see a significant improvement in the current ULs per k-bin. Specifically, each of our six most sensitive k-bins can *individually constrain* $kP_{\mathrm{CO}}(k) < 2400 - 4900\,\mu\mathrm{K}^2\mathrm{Mpc}^2$ at 95 % confidence. The maximum improvement between the two COMAP releases is around a factor 9 in the $k \sim 0.4\,\mathrm{Mpc}^{-1}$ bin. Note that the UL estimates are sensitive to both the uncertainty of a data point and its value. As the field-averaged FGPXS in the $k \sim 0.4\,\mathrm{Mpc}^{-1}$ bin is around -1.5σ below zero the resulting UL becomes the deepest even though according to Fig. 6 it is not the most sensitive k-bin.

When comparing COPSS to COMAP in Fig. 7 we see that where COMAP and COPSS have overlapping areas of high sensitivity, at $k < 0.8\,\mathrm{Mpc}^{-1}$, our 95 % ULs are significantly lower than those derived from the COPSS data points. This reflects the increased sensitivity of the COMAP FGPXS estimate already observed in Fig. 6. While none of the updated 95 % ULs are touching any of the two included models, a significantly larger region of the power spectrum space is excluded compared to only using the COPSS and COMAP ES limits, and our 95 % ULs are starting to encroach on the models that are not already excluded, including the fiducial model (Chung et al. 2022). Given our demonstrated ability to control systematic effects, and the constraints already achieved, detection of a CO power spectrum close to the fiducial model is within reach with further observations.

To conclude the discussion of the power spectrum results, the current COMAP power spectrum is the state-of-the-art CO LIM power spectrum dataset with around an order of magnitude more sensitivity and comparatively lower ULs at 95 % confidence than COPSS and COMAP ES, the only comparable CO(1–0) line-intensity mapping datasets in the literature. The presented power spectrum data points and resulting 95 % ULs further exclude a significant portion of the parameter space of possible CO models and provide the current best direct 3D constraints on the CO(1–0) power spectrum in the literature (Kovetz et al. 2017, 2019; Bernal & Kovetz 2022).

6. Null test results

As described in Sect. 3 we performed a set of null tests by computing the average cross-elevation FGPXS of a set of difference maps. All null tests were performed with the same pipeline and data selection as the power spectrum data shown in Sect. 5. The differencing variables chosen for the null tests were selected to test for correlations owing to a variety of potential systematic effects, e.g., environmental effects like weather, sidelobe pickup and pipeline diagnostics. In Table C.2 we show an overview of the selected null variables.

In total, 312 effective null tests were performed: 26 null test variables across three fields, cylindrical- and spherical-averaged FGPXS as well as separate tests for fast and slow azimuth data respectively. All of these can have different associated systematic effects. For instance, given that the telescope's scanning speed was changed to a lower azimuthal speed in May 2022, the fast and slow azimuth data (May 2022 – November 2023) may have very different mechanical vibrations that could cause spurious patterns in the maps (see Lunde et al. 2024, for examples).

For each of the effective null tests we calculate corresponding χ^2 probabilities-to-exceed (PTE), as described in Sect. 3. We provide a detailed list of these in Appendix C (see Table C.1).

Of the 312 null tests that we performed, the two lowest PTEs were found to be $\sim$ 0.6 %, which amounts to a random binomial probability of 27 %. Two of the null test χ^2-values were slightly outside the RND simulated χ^2-distribution and we therefore only have a lower limit of 99.5 % on their PTEs (because the numerical resolution of the simulation-based approach is $1/183 \approx 0.5$ %); this could be improved somewhat by using more RND realizations.

The PTEs are expected to follow a uniform distribution. As a consistency check, we therefore consider the PTE distributions of the performed null tests. In Fig. 8 we show the combined PTE-distribution for all separately performed null test (the corresponding distributions for each separately performed category of null tests can be seen in Fig. C.1 of Appendix C). To further gauge the uniformity of the histograms a Kolmogorov-Smirnov (KS) test was performed to see if the null test χ^2 PTEs were consistent with the null hypothesis of being drawn from a uniform distribution. The KS-test PTE-values are found in Table 2 (and also in the bottom row of Table C.1 of null test χ^2 PTEs in Appendix C). The lowest KS-test PTE of 5.5 %, corresponds to a binomial probability of around 35 % for the 12 performed KS tests. The maximum KS-test PTE is around 79 %, and the uniformity the uniformity of the entire set of PTEs is at the 58.7 % level.

We can therefore conclude that all the null tests and PTE uniformity tests have been passed and are consistent with the expected instrumental noise. As we do not claim any detection at this stage, the number and type of null tests performed are more than enough to ensure a sufficient data quality for our upper limits.

7. Conclusion

We have presented updated constraints on the cosmological CO(1–0) power spectrum at $2.4 < z < 3.4$, derived from the latest COMAP observations. These measurements are based on a

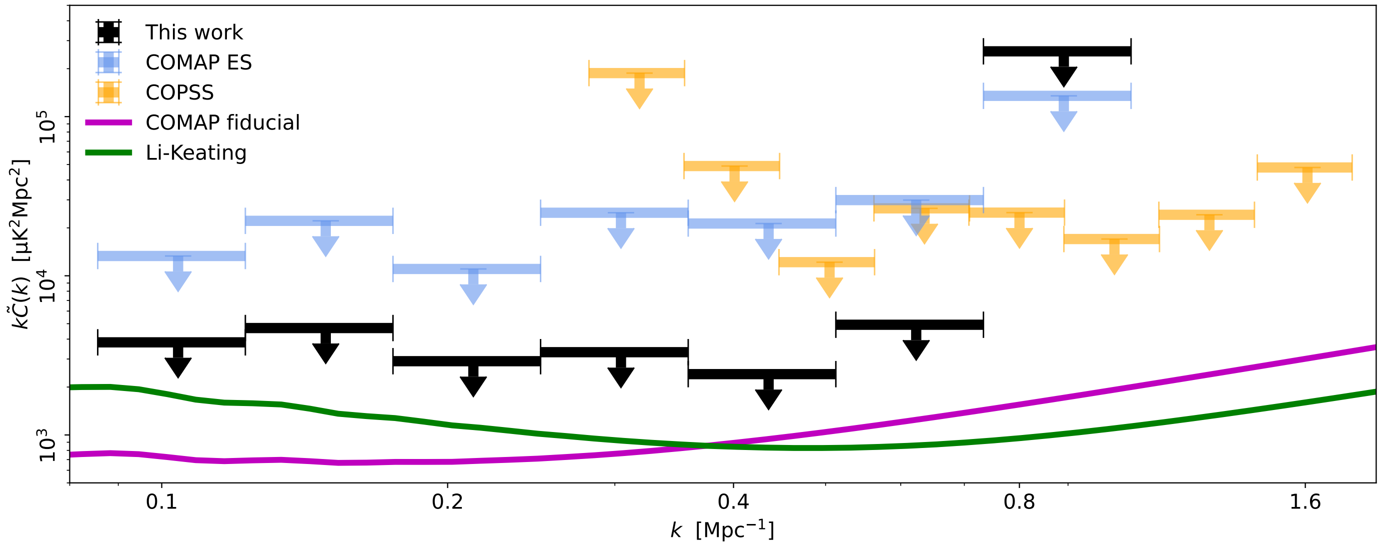


Fig. 7. Comparison of upper 95 % confidence limits (ULs) on the CO power spectrum as derived from the new COMAP data set (black), the COMAP ES analysis (blue; Ihle et al. 2022), and from COPSS (orange; Keating et al. 2016). The corresponding data points for each bin are shown in Fig. 5, and all ULs are derived using a positivity prior. The theoretical model predictions indicated by green and purple lines are the same as in Fig. 5. Note that because the data point of the FGPXS and COMAP ES centered at $k = 1.27\,\mathrm{Mpc}^{-1}$ in Figs. 4 and 5 have large uncertainties the corresponding 95 % UL are outside y-range of the figure.

Table 2. Kolmogorov-Smirnov uniformity test probabilities-to-exceed on the null test χ^2 probabilities-to-exceed.

Kolmogorov-Smirnov probabilities-to-exceed (KS PTEs) [%]												
	Spherically-averaged (1D)						Cylindrically-averaged (2D)					
	Field 1		Field 2		Field 3		Field 1		Field 2		Field 3	
Combined	Fast	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast	Slow
58.7	5.5	9.7	16.9	24.1	41.8	48.9	32.1	8.4	61.9	78.7	70.9	72.0

Notes. Probabilities-to-exceed of Kolmogorov-Smirnov (KS PTE) uniformity test of the null tests χ^2 PTEs (found in Table C.1 of Appendix C) in units percent of all three Fields, fast- and slow-moving azimuth scans (denoted as "Fast" and "Slow") as well as the spherically- and cylindrically-averaged FGPXS.

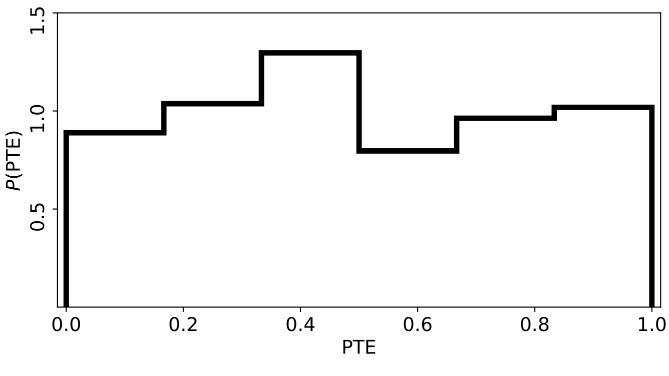


Fig. 8. Normalized distribution, $P(\mathrm{PTE})$, of χ^2 probabilities–to-exceed (PTEs) for all null tests performed on Field 1-3 combined. The PTE values corresponding to this histogram are found in Table C.1 in Appendix C. The Kolmogorov-Smirnov (KS) uniformity test on the samples contained in the illustrated distribution was found to yield a KS PTE of 58.7 % (see Table 2).

novel mean averaged feed-group pseudo cross-power spectrum (FGPXS) estimator, which is a slight modification of the feed-feed pseudo cross-power spectrum (FPXS) estimator used in the COMAP ES analysis (Ihle et al. 2022). The difference between these two estimators is that while the previous estimator evaluated cross-correlations between any two detector feeds, the new estimator evaluates cross-correlations between groups of feeds defined by common first downconversion (DCM1) local oscillators. The motivation for this is that feeds in these groups share some common instrumental systematic effects, and the new estimator is therefore more robust against such effects.

Quantitatively, all power spectrum bins were consistent with zero up to $\sim 2\sigma$, except for $k \sim 0.15\,\mathrm{Mpc}^{-1}$ which showed a 2–3σ excess in Fields 1 and 2; averaging over all three fields yields an excess of 2.7σ. Despite this single-bin excess, the total probabilities-to-exceed (PTE) with respect to a zero-signal model are 33.2 %, 19.5 % and 82.7 % for Fields 1–3, respectively, and 34 % when combining the data across fields. The resulting FGPXS spectrum derived from the latest COMAP data is thus statistically consistent with instrumental noise, and a detailed suite of null tests show no signs of residual systematic effects. At the same time, the slight excess at $k \sim 0.15\,\mathrm{Mpc}^{-1}$ is noteworthy; it could just be a regular noise fluctuation or the signature of some yet-to-be-discovered systematic effect. However, it could also be a small first hint of true cosmological CO fluctuations. More data are needed to determine its true nature.

Comparing with previous results, we find that the new COMAP power constraints are almost an order of magnitude stronger than the previous ES results (Ihle et al. 2022). In addition, when considering the power spectrum data points alone, the COPSS power spectrum (Keating et al. 2016) was found to be mostly consistent with the COMAP FGPXS, with only a mild $\sim 2.5\,\sigma$ tension in one of the bins. The volume-normalized sensitivity of the COMAP FGPXS was found to be around ten times that of the COPSS power spectrum estimate when comparing the respective most sensitive bins of the two experiments.

We developed a null test framework involving the difference between half-data maps that are split under variables believed to be associated with systematic effects. With the 26 split variables, three fields, the cylindrically- and spherically-averaged FGPXS as well as the fast- and slow-moving scans a total of 312 effective null tests were performed. Of these all passed within the expected instrumental uncertainties, ensuring the quality of our final data products.

To conclude, our power spectrum estimates and the resulting 95 % upper limits provide the most sensitive constraints on cosmic CO emission at $z \sim 2-3$ published to date and significantly reduce the allowed parameter space of possible CO emission models, the implications of which we explore further in the companion work of Chung et al. (2024). These results are a strong demonstration of COMAP's powerful capabilities and performance in terms of systematic effect mitigation, and the filtered data are still dominated by white noise even after three years of integration. Regular operations are still ongoing, and the data currently being gathered will put further pressure on possible CO emission models.

Acknowledgements. We acknowledge support from the Research Council of Norway through grants 251328 and 274990, and from the European Research Council (ERC) under the Horizon 2020 Research and Innovation Program (Grant agreement No. 772253 bits2cosmology and 819478 Cosmoglobe). This material is based upon work supported by the National Science Foundation under Grant Nos. 1517108, 1517288, 1517598, 1518282, 1910999, and 2206834, as well as by the Keck Institute for Space Studies under "The First Billion Years: A Technical Development Program for Spectral Line Observations". Parts of the work were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. PCB is supported by the James Arthur Postdoctoral Fellowship HP acknowledges support from the Swiss National Science Foundation via Ambizione Grant PZ00P2_179934. JK acknowledges support from a Robert A. Millikan Fellowship while at Caltech. DC is supported by a CITA/Dunlap Institute postdoctoral fellowship. The Dunlap Institute is funded through an endowment established by the David Dunlap family and the University of Toronto. Research in Canada is supported by NSERC and CIFAR. SEH acknowledges funding from an STFC Consolidated Grant (ST/P000649/1) and a UKSA grant (ST/Y005945/1) funding LiteBIRD foreground activities. JG acknowleges support from the Keck Institute for Space Science, NSF AST-1517108 and University of Miami and Hugh Medrano for assistance with cryostat design. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2024-00340759). NOS and JGSL extend a great thanks to Sigurd K. Næss for all the fruitful discussions in the office, and while biking through nature, during the last three years. This work was first presented at the Line Intensity Mapping 2024 conference held in Urbana, Illinois; we thank the organizers for their hospitality and the participants for useful discussions.

Software acknowledgments. Matplotlib for plotting (Hunter 2007); NumPy (Harris et al. 2020) and SciPy (Virtanen et al. 2020) for efficient numerics and array handling in Python; Astropy a community made core Python package for astronomy (Astropy Collaboration et al. 2013, 2018, 2022); Multi-node parallelization with MPI for Python (Dalcín et al. 2005, 2008; Dalcin et al. 2011; Dalcin & Fang 2021); Pixell (`https://github.com/simonsobs/pixell`) for handling sky maps in rectangular pixelization.

References

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Battye, R. A., Davies, R. D., & Weller, J. 2004, MNRAS, 355, 1339
Bernal, J. L. & Kovetz, E. D. 2022, A&A Rev., 30, 5
Bond, J. R. & Myers, S. T. 1996, ApJS, 103, 1
Breysse, P. C., Chung, D. T., Cleary, K. A., et al. 2022, ApJ, 933, 188
Breysse, P. C., Chung, D. T., & Ihle, H. T. 2023, MNRAS, 525, 1824
Breysse, P. C., Kovetz, E. D., Behroozi, P. S., Dai, L., & Kamionkowski, M. 2017, MNRAS, 467, 2996
Chung, D. T., Bangari, I., Breysse, P. C., et al. 2023, MNRAS, 520, 5305
Chung, D. T., Breysse, P. C., Cleary, K. A., et al. 2024, arXiv e-prints, arXiv:2406.07512
Chung, D. T., Breysse, P. C., Cleary, K. A., et al. 2022, ApJ, 933, 186
Chung, D. T., Breysse, P. C., Ihle, H. T., et al. 2021, ApJ, 923, 188
Cleary, K. A., Borowska, J., Breysse, P. C., et al. 2022, ApJ, 933, 182
Dalcin, L. & Fang, Y.-L. L. 2021, Computing in Science & Engineering, 23, 47
Dalcin, L. D., Paz, R. R., Kler, P. A., & Cosimo, A. 2011, Advances in Water Resources, 34, 1124
Dalcín, L., Paz, R., & Storti, M. 2005, Journal of Parallel and Distributed Computing, 65, 1108
Dalcín, L., Paz, R., Storti, M., & D'Elía, J. 2008, Journal of Parallel and Distributed Computing, 68, 655
Dunne, D. A., Cleary, K. A., Breysse, P. C., et al. 2024, ApJ, 965, 7
Foss, M. K., Ihle, H. T., Borowska, J., et al. 2022, ApJ, 933, 184
Gaunt, R. E. 2019, Statistica Neerlandica, 73, 176
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, ApJ, 567, 2
Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90
Ihle, H. T., Borowska, J., Cleary, K. A., et al. 2022, ApJ, 933, 185
Ihle, H. T., Chung, D., Stein, G., et al. 2019, ApJ, 871, 75
Keating, G. K., Marrone, D. P., Bower, G. C., & Keenan, R. P. 2020, ApJ, 901, 141
Keating, G. K., Marrone, D. P., Bower, G. C., et al. 2016, ApJ, 830, 34
Kovetz, E., Breysse, P. C., Lidz, A., et al. 2019, BAAS, 51, 101
Kovetz, E. D., Viero, M. P., Lidz, A., et al. 2017, arXiv e-prints, arXiv:1709.09066
Lamb, J. W., Cleary, K. A., Woody, D. P., et al. 2022, ApJ, 933, 183
Leung, J. S. Y., Hartley, J., Nagy, J. M., et al. 2022, ApJ, 928, 109
Li, T. Y., Wechsler, R. H., Devaraj, K., & Church, S. E. 2016, ApJ, 817, 169
Loeb, A. & Wyithe, J. S. B. 2008, Phys. Rev. Lett., 100, 161301
Lunde, J. G. S., Stutzer, N. O., Breysse, P. C., et al. 2024, arXiv e-prints, arXiv:2406.07510
Madau, P. & Dickinson, M. 2014, ARA&A, 52, 415
Madau, P., Meiksin, A., & Rees, M. J. 1997, ApJ, 475, 429
Nadarajah, S. & Pogány, T. K. 2016, Comptes Rendus. Mathématique, 354, 201
Peterson, J. B., Bandura, K., & Pen, U. L. 2006, arXiv e-prints, astro–ph/0606104
Rennie, T. J., Harper, S. E., Dickinson, C., et al. 2022, ApJ, 933, 187
Stein, G., Alvarez, M. A., & Bond, J. R. 2019, MNRAS, 483, 2236
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
Watts, D. J., Addison, G. E., Bennett, C. L., & Weiland, J. L. 2020, ApJ, 889, 130

Appendix A: Normalized CO power spectrum sensitivity

Table A.1 provides a comparison of the volume-normalized CO power spectrum sensitivity for the new COMAP data set with those power spectra derived from COMAP ES and COPSS; these are visualized in Fig. 6. The volume-normalized sensitivity is defined as

$$\xi_k = \sigma_k \sqrt{\Delta V_k}, \tag{A.1}$$

where σ_k is the uncertainty of a spherically-averaged power spectrum bin k with shell volume ΔV_k. This definition eliminates the effect of within-bin variance at each k-bin and provides a volume-independent measure that may be used to compare sensitivities between non-overlapping power spectrum bins and surveys. We see that the current COMAP power spectrum constraints reach a maximum sensitivity of one order of magnitude higher than the most sensitive COPSS (Keating et al. 2016) and COMAP ES (Ihle et al. 2022) bins. In addition, it is important to note that the regimes of maximum sensitivity differ between COMAP and COPSS, and this is due to their different instrumental designs and effective angular resolutions; COMAP is more sensitive in the large-scale clustering regime, while COPSS is more sensitive in the small-scale shot-noise regime.

Appendix B: Power spectrum data point values

For the interested reader we provide a list of power spectrum values and uncertainties, $k\tilde{C}(k)$ and $k\sigma_{\tilde{C}(k)}$ respectively, of the spherically- and field-averaged FGPXS data points seen in Fig. 5. These can be found in Table B.1.

Appendix C: Null test probabilities-to-exceed

In the following, we present a summary of the χ^2 probabilities-to-exceed (PTE) for each of our effective 312 null tests performed. The PTEs are found in Table C.1 and each null variable, and the corresponding acronyms are explained in detail in Table C.2.

Table C.1 is structured as follows: each row shows a different null variable in which the data was split in two, e.g. ambient temperature (ambt) or right and left moving azimuth sweeps (azdr). The columns are grouped into a hierarchical structure, as we performed null tests separately on spherically- and cylindrically-averaged FGPXS, for each field (Fields 1-3) as well as for data that was gathered before and after May 2022 when the scanning speed of the telescope was reduced. That is, because the fast- and slow-moving azimuth scans may have different associated systematic effects from, for example, mechanical vibrations in the telescope.

We present the distributions of PTEs of each separately performed null test in Fig. C.1 (see Fig. 8 in Sect. 6 for distribution of all null test PTEs considered jointly). As the distribution of PTEs is expected to be uniform we also performed a Kolmogorov-Smirnov (KS) test to find how probable it is that the PTE samples are drawn from a uniform distribution. These are shown for each separate null test category in the very last row of Table C.1 (and also in Table 2 in Sect. 6). For a discussion on the null test results see Sect. 6.

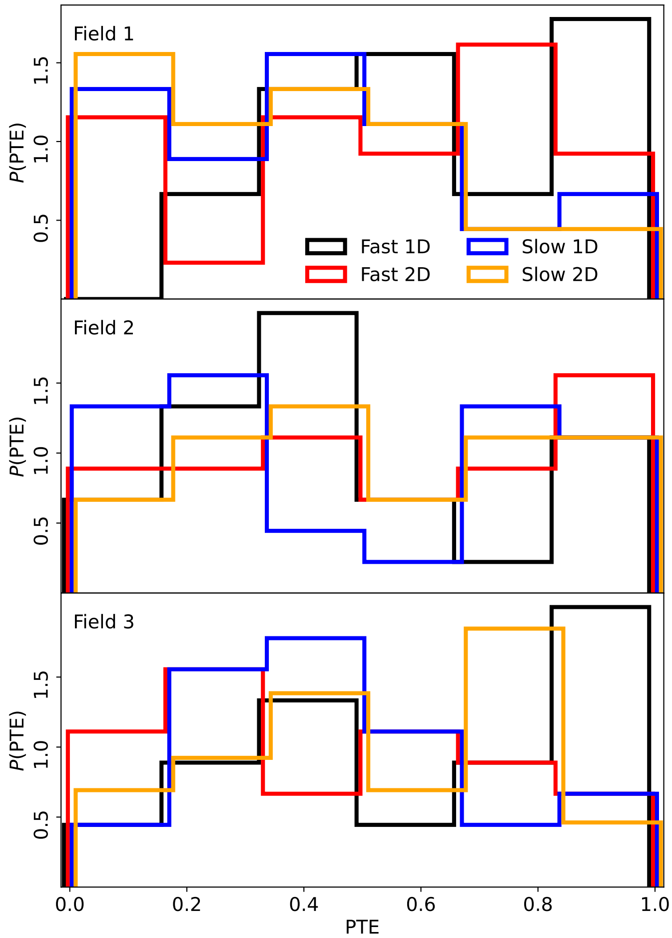


Fig. C.1. Normalized distribution, $P(\mathrm{PTE})$, of χ^2 probabilities–to-exceed (PTEs), separately, for all null tests performed on Field 1-3. The PTE values corresponding to this histogram are found in Table C.1. The Kolmogorov-Smirnov (KS) uniformity test results for the χ^2-samples of each of the separate PTE distributions can be found in Table 2 or the bottom row of Table C.1. The individual histograms are slightly offset w.r.t. each other for increased readability.

Appendix D: A simple end-to-end signal injection test

The same type of simulations used to estimate the filter transfer function (see Sect. 4.1) can also be used to perform a rudimentary end-to-end signal injection test to confirm the detectability of signal given our transfer function estimate. The signal injection pipeline is explained in more detail in Sect. 6.1 of Lunde et al. (2024), and we here limit the scope to the application thereof.

Appendix D.1: The signal injection tests

As the signal and noise in the raw data are affected by both the low-level analysis and instrumental effects described by Lunde et al. (2024), an important question to answer is whether we would be able to reconstruct the amplitude of an amplified CO signal within the estimated uncertainties using the earlier described FGPXS method. For instance, we know that several of the PCA filters in the low-level pipeline detailed by Lunde et al. (2024) can in principle act non-linearly if the CO-SNR becomes too high. One therefore has to verify that the filters remove equal amounts of signal at any given k-mode of the map in the filter

Table A.1. Normalized sensitivity in each COMAP and COPSS bin

Survey	k-center [Mpc^{-1}]	k-min [Mpc^{-1}]	k-max [Mpc^{-1}]	normalized sensitivity ξ_k [$10^3\mu\mathrm{K}^2\mathrm{Mpc}^{3/2}$]	$\xi^{\mathrm{min}}_{\mathrm{COPSS}}/\xi_k$
This work	0.1	0.09	0.12	1.25	7.3
	0.15	0.12	0.18	0.89	10.2
	0.21	0.18	0.25	1.24	7.3
	0.3	0.25	0.36	1.47	6.2
	0.44	0.36	0.51	2.6	3.5
	0.62	0.51	0.73	5.42	1.7
	0.89	0.73	1.05	181.92	0.0499
	1.27	1.05	1.5	9.3×10^6	9.7×10^{-7}
COMAP ES	0.1	0.09	0.12	6.47	1.4
	0.15	0.12	0.18	7.32	1.2
	0.21	0.18	0.25	9.16	1.0
	0.3	0.25	0.36	12.41	0.7
	0.44	0.36	0.51	21.28	0.4
	0.62	0.51	0.73	39.6	0.2
	0.89	0.73	1.05	91.5	0.0993
	1.27	1.05	1.5	28000	0.0003
COPSS bins	0.4	0.36	0.45	23.27	0.39
	0.5	0.45	0.57	12.04	0.754
	0.64	0.57	0.71	9.08	1.0
	0.8	0.71	0.9	9.16	0.991
	1.01	0.9	1.13	12.66	0.718
	1.27	1.13	1.42	23.16	0.392
	1.61	1.42	1.79	51.08	0.178
	2.02	1.79	2.84	277.81	0.033
	3.2	2.84	3.57	2060.78	0.004

Notes. Volume-normalized sensitivity, ξ_k, of each of our field-averaged power spectrum bins as well as the COPSS measurement (Keating et al. 2016). The normalized sensitivity ratio of COMAP (i.e. this work), COMAP ES (Ihle et al. 2022) and the individual COPSS bins relative to the most sensitive COPSS bin (Keating et al. 2016) (also seen in Fig. 7 as an orange marker) is given by $\xi^{\mathrm{min}}_{\mathrm{COPSS}}/\xi_k$.

Table B.1. Overview of FGPXS bin values and uncertainties.

k-center [Mpc^{-1}]	$k\tilde{C}(k)$ [$10^3\mu\mathrm{K}^2\mathrm{Mpc}^2$]	$k\sigma_{\tilde{C}(k)}$ [$10^3\mu\mathrm{K}^2\mathrm{Mpc}^2$]
0.1............	0.36	1.82
0.15	2.9	1.09
0.21	0.59	1.27
0.3............	1.19	1.26
0.44	−2.37	1.86
0.62	−2.48	3.24
0.89	101.5	90.9
1.27	-5.05×10^5	3.9×10^6

Notes. Bin values and uncertainties (respectively in the last two columns) of the spherically- and field-averaged FGPXS corresponding to our data points seen in Fig. 5.

transfer function estimation and for the actual signal estimation from the data. Otherwise, the final signal estimate computed obtained from the data would be biased.

We generate mock signal maps to use in this injection mechanism by applying the fiducial halo model to dark matter halos simulated with the peak-patch technique (Bond & Myers 1996; Stein et al. 2019). Furthermore, we use a raw COMAP data volume corresponding to all the fast-moving azimuth scans of Field 3. This roughly corresponds to the largest independently filtered data volume; i.e. the highest possible CO-SNR. We then boost the injected signal by a factor of three before injecting it into the raw time stream to ensure the signal is detectable above the instrumental noise. Subsequently, the TOD are filtered and binned into maps using the pipeline described by Lunde et al. (2024), before we compute the FGPXS signal estimate to see whether the injected signal was successfully recovered. Note, that only one signal realization is used because these high-realism mocks are expensive to produce. However, the test still functions as a simple qualitative "sanity check" that the pipeline works as intended. Future work will expand on this modest check by including further signal realizations and COMAP fields.

The resulting mock FGPXS data points as well as the auto-power spectrum of the input simulation can be seen in Fig. D.1. We can clearly see a high-significance excess that appears consistent the power spectrum of the input signal within the estimated error bars (which are estimated using the RND methodology as described in Sec. 2.3). The excess is large enough to place the computed χ^2-value of the mock data far outside the computed RND χ^2-distribution. Therefore, to assign a quantitative value to the significance of this mock detection, we instead use the simplified assumptions of approximately Gaussian uncertainties. When doing so, we obtain an estimated $\approx 6\sigma$ detection of non-zero power. Meanwhile, testing against the input signal we get a 1.5σ significance away from the model, meaning we recover the input signal within at most mild tension.

This exercise demonstrates that we can recover the input signal within the experimental errors, indicating that our pipeline,

Table C.1. Detailed overview of null test χ^2 probabilities-to-exceed.

	χ^2 probabilities-to-exceed [%]											
	Spherically-averaged (1D)						**Cylindrically-averaged (2D)**					
	Field 1		**Field 2**		**Field 3**		**Field 1**		**Field 2**		**Field 3**	
Null variable	**Fast**	**Slow**	**Fast**	**Slow**	**Fast**	**Slow**	**Fast**	**Slow**	**Fast**	**Slow**	**Fast**	**Slow**
ambt	57	33	8	13	76	22	77	45	49	26	70	74
wind	43	17	46	29	83	40	77	7	45	26	17	87
wint	41	49	64	93	17	62	7	24	90	91	47	76
half	86	16	21	6	14	43	49	3	87	39	25	68
odde	86	48	38	91	38	66	54	14	9	91	95	43
dayn	61	87	86	77	29	11	73	7	76	79	3	79
dtmp	25	19	38	49	92	$\geq 99.5^a$	14	27	36	83	27	42
hmty	95	52	98	70	51	17	90	41	81	79	29	44
pres	52	92	32	23	35	76	9	43	2	99	61	20
wthr	69	79	67	37	38	28	37	15	45	62	63	68
sune	89	58	95	27	96	28	40	78	79	8	7	42
modi	34	46	16	27	72	48	83	51	88	42	90	37
sudi	91	55	16	27	97	94	26	24	20	96	44	80
tsys	26	42	33	93	17	28	9	88	61	44	54	93
fpoO	44	97	11	3	64	62	60	56	62	83	74	52
fpoI	39	45	44	73	49	3	62	74	44	89	53	25
apoO	50	72	42	92	86	44	60	98	73	27	17	0.6
apoI	68	58	87	71	67	25	$\geq 99.5^a$	18	1	8	76	21
spoO	74	17	95	76	24	61	81	38	92	57	97	69
spoI	51	9	37	27	70	37	96	58	57	13	11	11
npca	89	51	19	80	86	98	92	49	16	45	74	55
pcaa	58	30	64	60	97	19	35	67	27	42	32	24
s01f	61	42	33	16	44	37	88	3	89	24	60	80
fk1f	85	16	1	20	36	49	78	56	1	68	11	17
al1f	30	14	50	16	83	62	81	48	90	67	49	39
azdr	43	42	34	98	93	71	0.6	27	18	20	11	3
KS-test	5.5	9.7	16.9	24.1	41.8	48.9	32.1	8.4	61.9	78.7	70.9	72.0

Notes. Null test χ^2 probabilities-to-exceed (PTE) in units percent. All tabulated PTE values, for all three Fields, fast- and slow-moving azimuth scans (denoted as "Fast" and "Slow"), are numerically computed from the RND ensemble. The last row indicates the Kolmogorov-Smirnov (KS) uniformity test PTE. The KS uniformity PTE of the entire table of PTEs is 58.7 %.
[(a)] The χ^2-value of this null test was slightly outside the simulated RND χ^2-distribution and we hence only have a lower limit of 99.5 % on the numerical PTE as the numerical resolution of the simulated distribution is roughly $1/183 \sim 0.5$ % from the RND ensemble size.

the full transfer function, and error bar estimation work as expected.

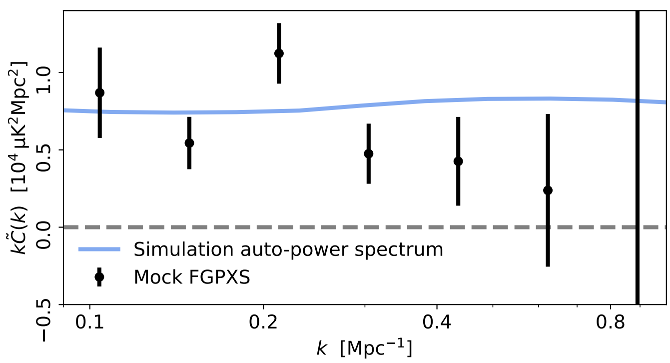


Fig. D.1. Example of the spherically-averaged FGPXS (black points) resulting from injecting a mock CO signal realization (blue input power spectrum) of the (line-broadened) COMAP fiducial model (Chung et al. 2021, 2022) with a boost factor of three into all fast-moving azimuth data of Field 3.

Table C.2. Detailed overview and explanation of null test variables

Null variable	Explanation
ambt	Ambient temperature at telescope site, as recorded by a nearby weather station.
wind	Wind speed, as recorded by a nearby weather station.
wint	Winter/summer split, the time difference to the middle of winter (15th of January).
half	Half-mission split, early versus late scans.
odde	Odd versus even scans.
dayn	Day/night split, time difference to 2 AM.
dtmp	Dew temperature, as recorded by a nearby weather station.
hmty	Humidity, as recorded by a nearby weather station.
pres	Air pressure, as recorded by a nearby weather station.
wthr	Bad weather and cloud coverage, predicted by a neural network trained on the raw data.
sune	Sun elevation.
modi	Average angular distance from the center of the field to the moon during the scan.
sudi	Average angular distance from the center of the field to the sun during the scan.
tsys	Average system temperature, as measured by the vane calibration, during the scan.
fpo0	$f_{\rm knee}$ value of a $1/f$ fit on the 0th-order $1/f$ gain fluctuation filter coefficient.[a]
fpo1	$f_{\rm knee}$ value of a $1/f$ fit on the 1st-order $1/f$ gain fluctuation filter coefficient.[a]
apo0	α value of a $1/f$ fit on the 0th-order $1/f$ gain fluctuation filter coefficient.[a]
apo1	α value of a $1/f$ fit on the 1st-order $1/f$ gain fluctuation filter coefficient.[a]
spo0	σ_0 value of a $1/f$ fit on the 0th-order $1/f$ gain fluctuation filter coefficient.[a]
spo1	σ_0 value of a $1/f$ fit on the 1th-order $1/f$ gain fluctuation filter coefficient.[a]
npca	Number of PCA components subtracted in the TOD filtering pipeline.
pcaa	Average amplitude of the fitted PCA components in the TOD filtering pipeline.
s01f	σ_0 value of a $1/f$ fit on the sideband-averaged time-domain data.
fk1f	$f_{\rm knee}$ value of a $1/f$ fit on the sideband-averaged time-domain data.
al1f	α value of a $1/f$ fit on the sideband-averaged time-domain data.
azdr	Scans split internally in left- vs right-moving pointing, in azimuth.

Notes. Explanation of the null test split variables. For all variables we show the abbreviation used in Table C.1 and a more detailed explanation of the null test variable
[(a)] As part of the TOD filtering a first order polynomial is fitted across the frequency bands for each time-sample. The 0th- and 1st-order polynomial components (as functions of time) tend to follow a $1/f$ spectrum, and a fit is performed on their TOD power spectra. See Lunde et al. (2024); Foss et al. (2022) for details.

Paper V

COMAP Pathfinder – Season 2 results III. Implications for cosmic molecular gas content at "Cosmic Half-past Eleven"

Astronomy & Astrophysics manuscript no. aanda
June 18, 2024

COMAP Pathfinder – Season 2 results
III. Implications for cosmic molecular gas content at "Cosmic Half-past Eleven"

D. T. Chung[1,2,3,⋆], P. C. Breysse[4,5], K. A. Cleary[6], D. A. Dunne[6], J. G. S. Lunde[7], H. Padmanabhan[8], N.-O. Stutzer[7], D. Tolgay[1,9], J. R. Bond[1,9,10], S. E. Church[11], H. K. Eriksen[7], T. Gaier[12], J. O. Gundersen[13], S. E. Harper[14], A. I. Harris[15], R. Hobbs[16], H. T. Ihle[7], J. Kim[17], J. W. Lamb[16], C. R. Lawrence[12], N. Murray[1,9,10], T. J. Pearson[6], L. Philip[18], A. C. S. Readhead[6], T. J. Rennie[14,19], I. K. Wehus[7], and D. P. Woody[16] (COMAP Collaboration)

[1] Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada
[2] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[3] Department of Astronomy, Cornell University, Ithaca, NY 14853, USA
[4] Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA
[5] Department of Physics, Southern Methodist University, Dallas, TX 75275, USA
[6] California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
[7] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, N-0315 Oslo, Norway
[8] Departement de Physique Théorique, Universite de Genève, 24 Quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
[9] Department of Physics, University of Toronto, 60 St. George Street, Toronto, ON, M5S 1A7, Canada
[10] David A. Dunlap Department of Astronomy, University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada
[11] Kavli Institute for Particle Astrophysics and Cosmology & Physics Department, Stanford University, Stanford, CA 94305, USA
[12] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[13] Department of Physics, University of Miami, 1320 Campo Sano Avenue, Coral Gables, FL 33146, USA
[14] Jodrell Bank Centre for Astrophysics, Department of Physics & Astronomy, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[15] Department of Astronomy, University of Maryland, College Park, MD 20742, USA
[16] Owens Valley Radio Observatory, California Institute of Technology, Big Pine, CA 93513, USA
[17] Department of Physics, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
[18] Brookhaven National Laboratory, Upton, NY 11973-5000
[19] Department of Physics and Astronomy, University of British Columbia, Vancouver, BC Canada V6T 1Z1, Canada

Received DD MMM YYYY; accepted DD MMM YYYY

ABSTRACT

The Carbon monOxide Mapping Array Project (COMAP) Pathfinder survey continues to demonstrate the feasibility of line-intensity mapping using high-redshift carbon monoxide (CO) line emission traced at cosmological scales. The latest COMAP Pathfinder power spectrum analysis is based on observations through the end of Season 2, covering the first three years of Pathfinder operations. We use our latest constraints on the CO(1–0) line-intensity power spectrum at $z \sim 3$ to update corresponding constraints on the cosmological clustering of CO line emission and thus the cosmic molecular gas content at a key epoch of galaxy assembly. We first mirror the COMAP Early Science interpretation, considering how Season 2 results translate to limits on the shot noise power of CO fluctuations and the bias of CO emission as a tracer of the underlying dark matter distribution. The COMAP Season 2 results place the most stringent limits on the CO tracer bias to date, at $\langle Tb \rangle < 4.8\,\mu$K. These limits narrow the model space significantly compared to previous CO line-intensity mapping results while maintaining consistency with small-volume interferometric surveys of resolved line candidates. The results also express a weak preference for CO emission models used to guide fiducial forecasts from COMAP Early Science, including our data-driven priors. We also consider directly constraining a model of the halo–CO connection, and show qualitative hints of capturing the total contribution of faint CO emitters through the improved sensitivity of COMAP data. With continued observations and matching improvements in analysis, the COMAP Pathfinder remains on track for a detection of cosmological clustering of CO emission.

Key words. galaxies: high-redshift – radio lines: galaxies – diffuse radiation

⋆ e-mail: dongwooc@cita.utoronto.ca

1. Introduction

Line-intensity mapping (LIM) surveys map the large-scale structure of the Universe in large cosmological volumes, but not through discrete resolved tracer sources. Rather, LIM surveys achieve this through unresolved emission in specific spectral lines, including lines associated with different phases of the star-forming interstellar medium (ISM) such as carbon monoxide (CO) and the [C ii] line from singly ionized carbon (see Kovetz et al. 2019 and Bernal & Kovetz 2022 for recent reviews). As part of a range of emerging interferometric and single-dish LIM surveys from radio to sub-millimeter wavelengths, the CO Mapping Array Project (COMAP; Cleary et al. 2022) is building a dedicated centimeter-wave LIM program to map the cosmic clustering of emission in the CO(1–0) and CO(2–1) lines from the epochs of galaxy assembly ($z \sim 3$, just before so-called "cosmic noon") and reionization ($z \sim 7$, "cosmic dawn"). COMAP science will encompass both the astrophysics of the assembly of molecular gas at these key epochs of galaxy evolution, and ultimately the cosmological implications of observed high-redshift large-scale structure traced by CO emission.

The first phase of COMAP is the COMAP Pathfinder, a 26–34 GHz spectrometer comprising a single-polarization 19-feed array of coherent receivers on a single 10-meter dish at the Owens Valley Radio Observatory (Lamb et al. 2022). The focus of the Pathfinder survey is on CO(1–0) emission from $z \sim 3$, or a lookback time of $\sim$ 11.5 Gyr. Around this "cosmic half-past eleven", we survey galaxies assembling towards the "cosmic noon" of peak cosmic star-formation activity (Somerville & Davé 2015; Förster Schreiber & Wuyts 2020). Following the Early Science analysis of Foss et al. (2022) and Ihle et al. (2022) based on the first season of observations (Season 1), the Season 2 data analysis by Lunde et al. (2024) and Stutzer et al. (2024) encompasses three years of observations and improved data cleaning methods for almost an order-of-magnitude increase in power spectrum sensitivity.

With such progress continuing to demonstrate the feasibility of CO LIM survey operations and low-level data analysis, we present here the corresponding update on our understanding of CO(1–0) emission at $z \sim 3$. We carry out a high-level analysis of the power spectrum constraints of Stutzer et al. (2024) to answer the following questions:

– How much does the increased data volume improve constraints on the clustering and shot noise power of cosmological CO(1–0) emission at $z \sim 3$?
– Can COMAP Season 2 data better constrain the empirical connection between CO emission and the underlying structures of dark matter?

We structure the paper as follows. In Sect. 2 we outline our methodology for interpretation, including but no longer limited to methods previously used in Chung et al. (2022). We discuss the results of our analysis in Sect. 3, and implications for understanding CO emission and interpreting past and future CO LIM surveys in Sect. 4. We end with our primary conclusions and future outlook in Sect. 5.

We assume a ΛCDM cosmology with parameters $\Omega_m = 0.286$, $\Omega_\Lambda = 0.714$, $\Omega_b = 0.047$, $H_0 = 100h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ with $h = 0.7$, $\sigma_8 = 0.82$, and $n_s = 0.96$, to maintain consistency with previous COMAP simulations (starting with Li et al. 2016). Distances carry an implicit h^{-1} dependence throughout, which propagates through masses (all based on virial halo masses, proportional to h^{-1}) and volume densities ($\propto h^3$). Logarithms are base-10 unless stated otherwise.

2. Methods

The primary target of the COMAP Pathfinder is the power spectrum of spatial-spectral emission after subtraction of continuum emission and systematic effects. Any residual fluctuations should predominantly arise from clustered populations of CO-emitting high-redshift galaxies, meaning that we interpret any constraints on the residual emission as constraints on these CO emitters. In the simplest possible model, the power spectrum as a function of comoving wavenumber k consists of the matter power spectrum $P_m(k)$ scaled by some amplitude A_{clust}, plus a scale-independent shot noise amplitude P_{shot}:

$$P(k) = A_{\mathrm{clust}} P_m(k) + P_{\mathrm{shot}}. \quad (1)$$

The matter power spectrum describes the distribution of matter density contrast across comoving space, and evolves with redshift as large-scale structure forms and grows. The spatial-spectral fluctuations in CO brightness temperature across cosmological scales trace the clustering of the underlying matter fluctuations with some bias, which informs the clustering amplitude A_{clust}. In combination with halo models of CO emission that postulate average CO luminosities per halo of collapsed dark matter, constraining A_{clust} (or related quantities) and P_{shot} allows us to understand not just the global cosmic abundance of CO, but also the relative contribution of different sizes of halos and thus of galaxies. Estimation of the CO line power spectrum $P(k)$ is thus a key target of COMAP low-level analyses.

The goal of this section is to outline methods for the kind of analyses suitable for the current level of sensitivity achieved by the COMAP dataset. First, Sect. 2.1 reviews the COMAP Season 2 power spectrum results in relation to previous work. Then, Sect. 2.2 reviews a simple two-parameter analysis as carried out by Chung et al. (2022) for COMAP Early Science, constraining the clustering and shot noise amplitudes and only indirectly using halo models to support physical interpretation. Finally, Sect. 2.3 outlines a five-parameter analysis to directly constrain a halo model of CO emission, as carried out by Chung et al. (2022) to derive priors for COMAP Early Science but incorporating COMAP data for the first time.

2.1. Foundational data: COMAP Season 2 power spectrum constraints

The present work makes use of the results of Stutzer et al. (2024), which derived updated power spectrum constraints based on COMAP Pathfinder survey data collected across 17500 hours over three fields of 2–3 deg^2 each, between its commissioning in May 2019 and the end of the second observing season in November 2023. We also make use of the prior work of the CO Power Spectrum Survey (COPSS; Keating et al. 2016), which performed a pilot CO LIM survey targeting largely the same observing frequencies, but with an interferometric dataset probing smaller scales. The COMAP observations are subject to the effects of instrument and pipeline response, such as filtering, pixelization, and beam smoothing. However, the results as considered in this work correct for these effects by applying the inverse of the estimated power spectrum transfer function per k-bin. We expect mode mixing in COMAP data is still at the level of Ihle et al. (2022) at most, that is to say less than 20% for the comoving wavenumber range of $k \gtrsim 0.1\,\mathrm{Mpc}^{-1}$ that we consider.

Fig. 1 shows these results alongside the range of expectations for the $z \sim 3$ CO(1–0) emission power spectrum from empirical modeling in the decade leading up to this dataset (Pullen et al. 2013; Li et al. 2016; Padmanabhan 2018; Keating et al.

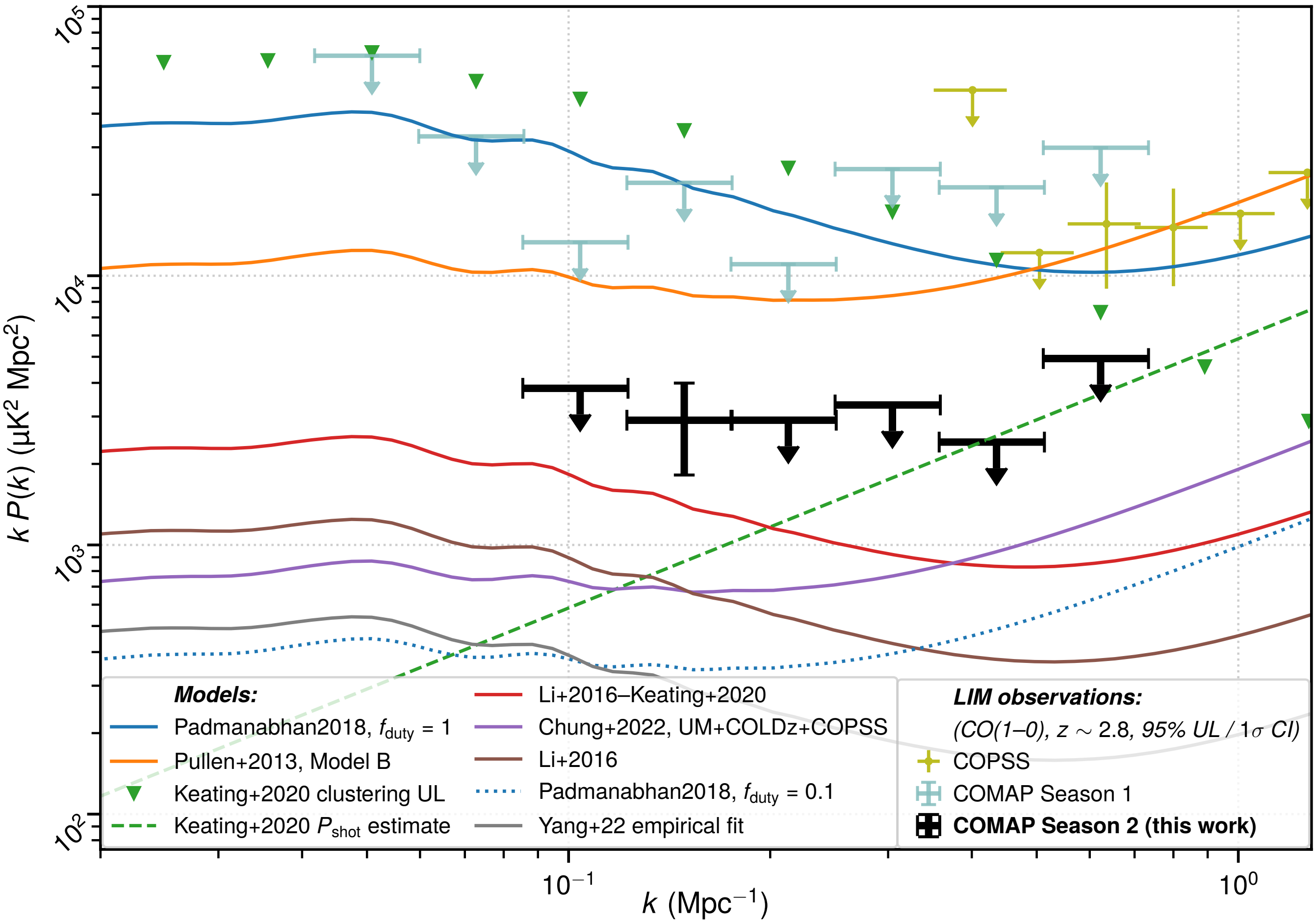


Fig. 1. COMAP Season 2 95% upper limits (given $P(k) > 0$) on the $z \sim 3$ CO(1–0) power spectrum, with analogous limits from COPSS (Keating et al. 2016) and COMAP Season 1 (Chung et al. 2022). Some k-bins in COPSS and COMAP Season 2 data show marginal excesses, influencing analyses in this work; we thus show 1σ intervals for these bins unlike in Stutzer et al. (2024). We also show predictions based on Chung et al. (2022), Padmanabhan (2018), Pullen et al. (2013), Li et al. (2016), and Yang et al. (2022), plus a variation on the Li et al. (2016) model from Keating et al. (2020), and the Keating et al. (2020) re-analysis of COPSS constraining clustering (triangles) and shot-noise amplitudes (dashed line).

2020; Chung et al. 2022; Yang et al. 2022). These models either postulate a connection between dark matter halo properties and CO luminosity via intermediate galaxy properties like star-formation rate (SFR), or directly model the halo–CO connection constrained by observed CO luminosity functions and CO LIM measurements.

Of the models shown in Fig. 1, only the models of Padmanabhan (2018) and Chung et al. (2022) fall into the latter category. Keating et al. (2020) also provide empirical estimates for the clustering and shot noise amplitudes, but this is simply based on decomposing the COPSS measurement of Keating et al. (2016) into clustering and shot noise components, rather than a detailed halo model. In a different context Keating et al. (2020) do provide a halo model, which we term the Li et al. (2016)–Keating et al. (2020) model, varying the Li et al. (2016) model by using the same halo–SFR connection from Behroozi et al. (2013a,b) but replacing the SFR–CO connection (via infrared luminosity) derived from a compilation of local and high-redshift galaxies (Carilli & Walter 2013) with one based on a local sample observed by Kamenetzky et al. (2016).

Even before any detailed analyses, compared to COMAP Season 1 we clearly see an increasing rejection of Model B of Pullen et al. (2013) and of the Padmanabhan (2018) model with CO emission duty cycle $f_{\rm duty} = 1$. We refer the reader to the Early Science work of Chung et al. (2022) for the implications of excluding these models. As with COMAP Early Science, we exclude these models in the clustering regime, rather than the shot-noise dominated scales surveyed by COPSS. However, the COMAP Season 2 sensitivity is sufficient to exclude these models clearly in *individual* k-bins of width $\Delta(\log k\,{\rm Mpc}) = 0.155$, rather than having to rely on a co-added measurement across all k as was necessary in Early Science. For reference, we show in Appendix A the original data points behind these upper limits, in a way that more closely resembles Fig. 4 of Stutzer et al. (2024).

Note also a weak tension against the previous positive COPSS measurement in overlapping k-ranges. The original COPSS analysis of Keating et al. (2016) measured the CO power spectrum at $k = 1h\,{\rm Mpc}^{-1}$ to be $P(k) = (3.0 \pm 1.3) \times 10^3 h^{-3}\,\mu{\rm K}^2\,{\rm Mpc}^3$, for a best estimate of $P(k = 0.7\,{\rm Mpc}^{-1}) = 8.7 \times 10^3\,\mu{\rm K}^2\,{\rm Mpc}^3$. This is co-added across the entire k-range spanned by COPSS, with the highest sensitivity achieved around $k = 0.5h$–$2h\,{\rm Mpc}^{-1}$. By contrast, in a single k-bin spanning $k =$ 0.52–0.75 Mpc^{-1}, the present COMAP data places a 95% upper limit of $7.9 \times 10^3\,\mu{\rm K}^2\,{\rm Mpc}^3$, lying below the COPSS co-added best estimate. However, the COPSS result is itself only a tentative one at $\approx 2.3\sigma$ significance, and so there is no statistically significant discrepancy. COMAP data are also entirely consistent with the estimate of $P_{\rm shot} = 2.0^{+1.1}_{-1.2} \times 10^3 h^{-3}\,\mu{\rm K}^2\,{\rm Mpc}^3$ from the later re-analysis of COPSS data by Keating et al. (2020), which marginalized over the possible contribution to $P(k)$ from clustering. In fact our power spectrum results show a marginal excess at $k \approx 0.15\,\mathrm{Mpc}^{-1}$ that, while well below the upper limit implied by the direct COPSS re-analysis of Keating et al. (2020), does tentatively indicate a preference for models like the Li et al. (2016)–Keating et al. (2020) model. The remainder of this work will establish this preference more quantitatively, and consider other implications of these results.

2.2. Two-parameter analysis: Constraining CO tracer bias and shot noise

The most direct way to analyze the COMAP Season 2 constraints is to decompose the CO power spectrum into clustering and shot noise terms as in Eq. (1), with a fixed cosmological model and no assumptions around detailed astrophysical modeling. The COMAP data then constrain the possible range of values for $A_{\rm clust}$ and $P_{\rm shot}$, which we may then compare to model predictions for these amplitudes for the clustering and shot noise contributions to the power spectrum.

However, physical interpretation requires some amount of guidance from models. Consider a halo model of CO emission where halos of virial mass M_h emit with CO luminosity $L(M_h)$. Suppose that we know the halo mass function dn/dM_h describing the differential number density of halos of mass M_h, and the bias $b_h(M_h)$ with which the halo number density contrast traces the continuous dark matter density contrast. Then the cosmological fluctuations in CO(1–0) line temperature trace the underlying dark matter fluctuations with a linear scaling of

$$\langle Tb\rangle \propto \int dM_h\,\frac{dn}{dM_h}L(M_h)b_h(M_h). \tag{2}$$

This should be understood as a mean line temperature–bias product, with appropriate normalization factors applied to convert luminosity density to brightness temperature. We may also ascribe a dimensionless bias b to CO emission contrast by dividing out the average line temperature or luminosity density:

$$b = \frac{\int dM_h\,(dn/dM_h)\,L(M_h)b_h(M_h)}{\int dM_h\,(dn/dM_h)\,L(M_h)}. \tag{3}$$

Furthermore, any halo model of $L(M_h)$ will predict the shot noise, proportional to the second bias- and abundance-weighted moment of the $L(M_h)$ function rather than the first moment:

$$P_{\rm shot} \propto \int dM_h\,\frac{dn}{dM_h}L^2(M_h)b_h(M_h). \tag{4}$$

The quantity $P_{\rm shot}$ directly describes the shot noise amplitude, but the same is not true of $\langle Tb\rangle$ in relation to the clustering amplitude. In real comoving space we would expect $A_{\rm clust} = \langle Tb\rangle^2$, but redshift-space distortions (RSD) enhance the clustering term as large-scale structure coherently attracts galaxies (Kaiser 1987; Hamilton 1998). In the linear regime of small k, and given that $\Omega_m(z) \approx 1$ at COMAP redshifts,

$$A_{\rm clust} \approx \langle Tb\rangle^2\left(1+\frac{2}{3b}+\frac{1}{5b^2}\right). \tag{5}$$

Based on prior modeling efforts, we consider $b > 2$ to be a fairly conservative lower bound on CO tracer bias, as outlined by Chung et al. (2022). This bound on b in turn allows us to bound $\langle T\rangle = \langle Tb\rangle/b$ based on an upper bound on $A_{\rm clust}$.

We consider two variants of a two-parameter analysis of the COMAP data, the same carried out in Chung et al. (2022).

1. The first variant is a model-agnostic evaluation of the likelihood of different values of $A_{\rm clust}$ and $P_{\rm shot}$ given the $P(k)$ data points available from the COPSS results of Keating et al. (2016) and/or from COMAP data through Season 2. We only invoke a conservative limit of $b > 2$ to obtain an upper bound on $\langle T\rangle$ from our constraint on $A_{\rm clust}$.
2. The second variant assumes that given values for $\langle Tb\rangle^2$ and $P_{\rm shot}$, we can expect specific values for b and for an effective line width $v_{\rm eff}$ describing the suppression of the high-k CO power spectrum from line broadening (Chung et al. 2021). Exploration of an empirically constrained model space informs fitting functions for b and $v_{\rm eff}$ given only $\langle Tb\rangle^2$ and $P_{\rm shot}$, as provided in Appendix B of Chung et al. (2022), which then enter into calculation of the redshift-space $P(k)$ accounting for RSD and line broadening. We can directly compare this $P(k)$ to our $P(k)$ data to evaluate the likelihood of different values of $\langle Tb\rangle^2$ and $P_{\rm shot}$. We refer to this variant as the "b- and $v_{\rm eff}$-informed" analysis, versus the first "b- and $v_{\rm eff}$-agnostic" version.

We may then compare likely and unlikely regions of this two-parameter space to model predictions.

2.3. Five-parameter analysis: Directly constraining the halo–CO connection

Neither variant of our two-parameter analysis truly directly constrains the physical picture of CO emission, only a clustering term and a shot noise term. Given a fixed set of power spectrum measurements, the two-parameter analysis will broadly project likelihood contours favouring either high clustering and low shot noise, or low clustering and high shot noise. Yet physical models should impose a strong prior such that clustering and shot noise co-vary, given that the shot noise also tracks with luminosity density, albeit at a higher order – cf. Eq. (4).

Directly modeling and constraining $L(M_h)$ thus has its uses. While dark matter halos are not themselves the direct source of CO emission or indeed any baryonic physics, a halo model of CO emission still serves as a simple way to physically ground interpretation of our CO measurements and introduce priors based on other empirical constraints on the galaxy–halo connection.

2.3.1. Parameterization and derivation of "UM+COLDz" posterior

To model $L(M_h)$, we use the same parameterization and data-driven procedure as in Chung et al. (2022). One of the datasets driving this procedure is provided by the CO Luminosity Density at High-z (COLDz) survey (Pavesi et al. 2018; Riechers et al. 2019), which identified line candidates at $z \sim 2.4$ through an untargeted interferometric search. In Chung et al. (2022) we also introduced somewhat informative priors based loosely on the work of Behroozi et al. (2019), which devised the UniverseMachine (UM) framework for an empirical model of the galaxy–halo connection by connecting halo accretion histories to a minimal model of stellar mass growth. We thus once again combine these "UM" priors with COLDz data and a basic $L(M_h)$ parameterization, just as in Chung et al. (2022).

We assume a double power law for the linear average $L(M_h)$. In observer units,

$$\frac{L'_{\rm CO}(M_h)}{\mathrm{K\,km\,s^{-1}\,pc^2}} = \frac{C}{(M_h/M)^A + (M_h/M)^B}. \tag{6}$$

For CO(1–0),

$$\frac{L_{\rm CO}(M_h)}{L_\odot} = 4.9\times10^{-5}\times\frac{L'_{\rm CO}(M_h)}{\rm K\,km\,s^{-1}\,pc^2}. \quad (7)$$

We also model stochasticity albeit in a highly simplistic fashion, assuming some level of log-normal scatter σ (in units of dex) about the average relation. We inherit this practice from the common use of log-normal distributions to model intrinsic scatter in, e.g., the halo–SFR connection (e.g.: Behroozi et al. 2013a,b) and the halo–CO connection as modeled for previous early COMAP forecasts (Li et al. 2016).

The somewhat informative "UM" priors for the five free parameters of $L(M_h)$ are as follows:

$$A = -1.66 \pm 2.33, \quad (8)$$
$$B = 0.04 \pm 1.26, \quad (9)$$
$$\log C = 10.25 \pm 5.29, \quad (10)$$
$$\log(M/M_\odot) = 12.41 \pm 1.77. \quad (11)$$

For log-normal scatter, we assume an initial prior of $\sigma = 0.4\pm0.2$ (dex), taking cues from Li et al. (2016) for the central value and slightly broadening the prior.

We then narrow these priors further by matching the luminosity function constraints of the COLDz survey. The matching procedure is similar to that used in Chung et al. (2022). However, that procedure used a snapshot from the Bolshoi–Planck simulation, used by (Behroozi et al. 2019) but slightly discrepant against our fiducial cosmology. Here, we use our own peak-patch mocks (Stein et al. 2019) with virial masses matched to the halo mass function of Tinker et al. (2008). We extract halos from $z \in (2.35, 2.45)$ (or $\chi \in (5720.37, 5844.19)$ Mpc) from these peak-patch mocks, since we are specifically trying to match a luminosity function constraint at $z \sim 2.4$. We thus obtain 161 independent realizations of a halo catalogue from a $1140\times1140\times124\,\text{Mpc}^3 = 0.16\ \text{Gpc}^3$ box, comparable to the Bolshoi–Planck snapshot with a box size of (250/0.678) Mpc (or a volume of 0.05 Gpc^3). A Markov chain Monte Carlo (MCMC) procedure identifies the posterior (narrowed prior) based on a likelihood calculation in addition to the mildly informative priors outlined above. At each step:

- We use the sampled $L(M_h)$ parameters to convert a random peak-patch realization of halo masses into CO luminosities.
- We then fit a Schechter function to the resulting CO luminosity function of that randomly chosen peak-patch box.
- We calculate the log-likelihood by comparing the Schechter function fit against the COLDz posterior for the Schechter function parameters via a kernel density estimator.

The MCMC uses 250 walkers for 4242 steps, and we discard the first 1000 steps as a burn-in phase.

The result is an informed distribution, which we call the "UM+COLDz" posterior, for the five parameters $\{A, B, C, M, \sigma\}$ of our $L(M_h)$ model.

2.3.2. Derivation of posteriors incorporating CO LIM data

To derive posteriors based on CO LIM power spectrum measurements, we rerun the same MCMC procedure used to derive the "UM+COLDz" distribution with additional contributions to the likelihood from any discrepancy with the CO LIM results. In other words, we introduce additive log-likelihood terms,

$$\Delta(\log\mathcal{L}) \propto -\sum_k \frac{[P_{\rm model}(k) - P_{\rm data}(k)]^2}{\sigma^2[P_{\rm data}(k)]}, \quad (12)$$

evaluated against each dataset $P_{\rm data}(k)$ with error $\sigma[P_{\rm data}(k)]$ for the model $P_{\rm model}(k)$ drawn at each MCMC step.[1]

Using our fiducial cosmology and the halo mass function of Tinker et al. (2008), we numerically evaluate closed-form expressions describing the CO power spectrum at $z \sim 2.8$. We evaluate the above log-likelihood terms against the predicted $P_{\rm model}(k)$ without imposing positivity priors, which would be redundant with the always positive predictions of our $P(k)$ model.

We consider three (combinations of) datasets:

- The "UM+COLDz+COPSS" posterior derives from considering only the addition of COPSS data points as shown in Keating et al. (2016).
- The "UM+COLDz+COPSS+COMAP S1" posterior derives from considering both COPSS and COMAP Early Science $P(k)$ constraints from Season 1 data.
- The "UM+COLDz+COPSS+COMAP S2" posterior derives from considering constraints from both COPSS and the present work on COMAP Pathfinder data through Season 2.

While the MCMC procedure itself evaluates posteriors for $\{A, B, C, M, \sigma\}$, we can use the resulting sampling of parameter space to obtain posterior distributions for derived quantities like $L(M_h)$, $\langle T\rangle$, b, and $P_{\rm shot}$, and we will look for how (if at all) the "UM+COLDz+COPSS+COMAP S2" posterior distinguishes itself from posteriors based on only previous data.

3. Results

Having outlined the datasets and methods used in the analyses, we now review the results in relation to previous models and results. We consider outcomes of the two-paramater analysis identifying overall amplitudes for clustering and shot noise power in Sect. 3.1, followed by outcomes of the five-parameter analysis fitting for the $L(M_h)$ relation in Sect. 3.2.

3.1. Two-parameter analysis

We summarize the results of the two-parameter analysis of the COMAP results in Table 1, and show in Fig. 2 the probability distributions when considering only COMAP data up to Season 2 ("COMAP S2" in Table 1). We find a factor of 5 improvement in our ability to constrain $P_{\rm shot}$ from above with COMAP data alone up to Season 2 compared to COMAP Early Science alone, and a factor of 2 improvement in upper limits for the clustering amplitude. In fact, framing sensitivity to clustering purely in terms of the upper limit achieved downplays our gain. Where the COMAP Early Science analysis effectively gave a maximum *a posteriori* (MAP) estimate of zero for $A_{\rm clust}$ and $\langle Tb\rangle^2$, Fig. 2 shows that the likelihood distributions peak at positive values of these parameters under COMAP Season 2 constraints.

We also show in Fig. 2 model predictions for $A_{\rm clust}$ and $P_{\rm shot}$, or for $\langle Tb\rangle^2$ and $P_{\rm shot}$. As expected, all models not shown to be excluded by the COMAP Season 2 data at 95% confidence in Fig. 1 are consistent to within 2σ of the MAP estimate from the COMAP Season 2 likelihood analysis, including the COMAP Early Science fiducial model from Chung et al. (2022). That said, the most favoured model (within 1σ of the MAP estimate) is the Li et al. (2016)–Keating et al. (2020) model used

[1] This approximates the likelihood as Gaussian and independent between k-bins, which we consider to be a reasonable approximation at least for COMAP Season 2 data. In obtaining the $P(k)$ result, Stutzer et al. (2024) found that on average, any single k-bin correlated with any other k-bin at a level of $\lesssim 10\%$.

Data	b- and $v_{\rm eff}$-agnostic: $A_{\rm clust}$ (μK^2)	$P_{\rm shot}/10^3$ (μK^2 Mpc^3)	b- and $v_{\rm eff}$-informed: $\langle Tb\rangle^2$ (μK^2)	$P_{\rm shot}/10^3$ (μK^2 Mpc^3)	b- and $v_{\rm eff}$-agnostic: $\langle T\rangle$ (μK)	$\rho_{\rm H2}/10^8$ ($M_\odot$ Mpc^{-3})	b- and $v_{\rm eff}$-informed: $\langle T\rangle$ (μK)	$\rho_{\rm H2}/10^8$ ($M_\odot$ Mpc^{-3})
COPSS	< 630	$5.7^{+4.2}_{-3.6}$	< 345	$12.1^{+7.5}_{-6.4}$	$< 11.$	< 7.4	< 9.3	< 6.4
COMAP S1	< 66	< 19	< 49	< 24	< 3.5	< 2.4	< 3.5	< 2.5
COMAP S1+COPSS	< 69	$6.8^{+3.8}_{-3.5}$	< 51	$11.9^{+6.8}_{-6.1}$	< 3.5	< 2.5	< 3.6	< 2.5
COMAP S2	**< 31**	**< 3.7**	**< 23**	**< 4.9**	**< 2.4**	**< 1.6**	**< 2.4**	**< 1.7**
COMAP S2+COPSS	< 30	< 4.8	< 23	< 6.1	< 2.3	< 1.6	< 2.4	< 1.7

Table 1. Results from two-parameter analyses of CO power spectrum measurements for clustering amplitude ($A_{\rm clust}$ or $\langle Tb\rangle^2$) and shot noise power ($P_{\rm shot}$), assuming any deviation from zero describes CO(1–0) emission at $z \sim 3$. For comparison, we also show results from using only COPSS data or COMAP data through Season 1; we indicate in bold type the results from using COMAP data through Season 2 (without COPSS data). We quote 68% intervals for $P_{\rm shot}$ in the "COPSS" and "COMAP S1+COPSS" analyses; otherwise we quote 95% upper limits.

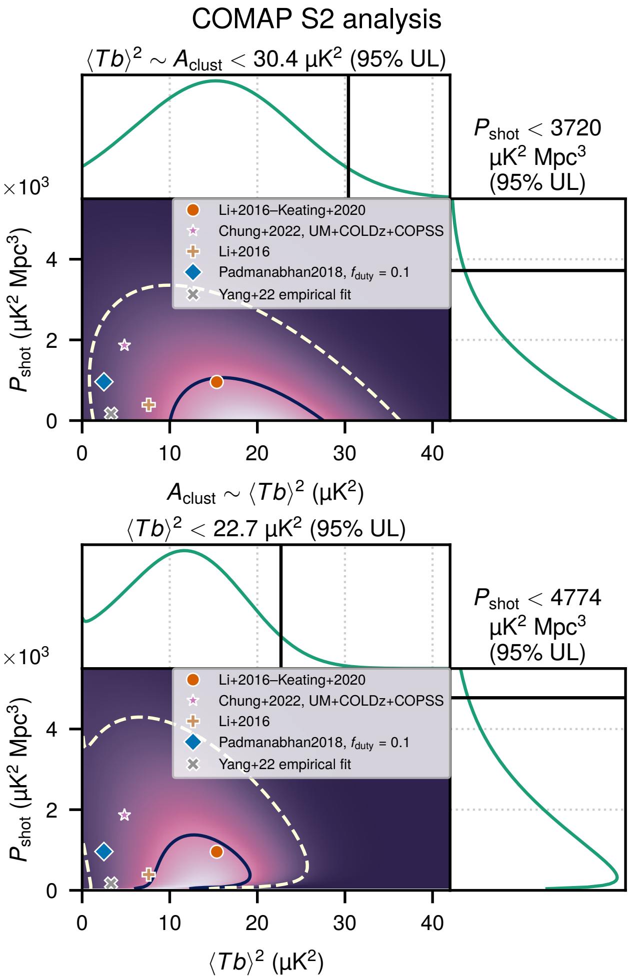


Fig. 2. Likelihood contours and marginalized probability distributions for the clustering and shot-noise amplitudes of the CO power spectrum, conditioned on COMAP Season 2 data, in b- and $v_{\rm eff}$-agnostic (upper) and -informed (lower) analyses. Black solid lines plotted with the 1D marginalized distributions indicate the 95% upper limits for each parameter. The solid and dashed 2D contours are meant to encompass 39% and 86% of the probability mass (delineated at $\Delta\chi^2 = \{1, 4\}$ relative to the minimum χ^2, corresponding to 1σ and 2σ for 2D Gaussians). We show the clustering and shot noise amplitudes for a subset of the models plotted in Fig. 1. Models shown in Fig. 1 but not shown here have values of $A_{\rm clust}$ or $\langle Tb\rangle^2$ well beyond the 2σ regions shown.

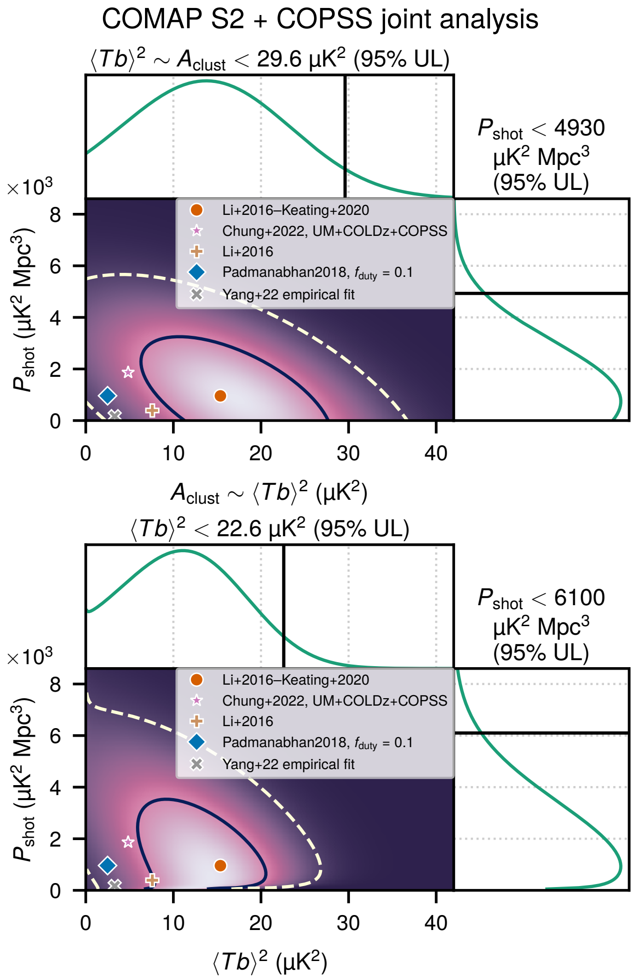


Fig. 3. Same as Fig. 2, but with a combination of COMAP Season 2 and COPSS data conditioning the likelihood.

to explain the results of the mm-wave Intensity Mapping Experiment (mmIME; Keating et al. 2020). This finding is consistent between the b- and $v_{\rm eff}$-agnostic and -informed analyses.

The resulting constraints on $\langle T\rangle$ given $b > 2$ are also consistent between these analyses to within a few percent. Going forward we will quote $\langle Tb\rangle < 4.8\,\mu$K or $\langle T\rangle < 2.4\,\mu$K, consistent with both of our "COMAP S2" standalone analyses as well as both of the "COMAP S2+COPSS" joint analyses as we show in Table 1. As in Chung et al. (2022) we can convert any estimate of $\langle T\rangle$ into an estimate for cosmic molecular gas density:

$$\rho_{\rm H2} = \alpha_{\rm CO}\,\langle T\rangle\,H(z)/(1+z)^2. \quad (13)$$

We show the resulting bounds on $\rho_{\rm H2}$ in Table 1 alongside the original bounds on $\langle T\rangle$, given $\alpha_{\rm CO} = 3.6\,M_\odot\,({\rm K\,km\,s^{-1}\,pc^2})^{-1}$ and the Hubble parameter $H(z)$ at the central COMAP redshift of $z \sim 2.8$. Although some works have advocated for values of $\alpha_{\rm CO}$ (Bolatto et al. 2013; Scoville et al. 2016) higher by as much as a factor of two, our chosen value follows the one most commonly used by previous CO line search and line-intensity mapping analyses (e.g.: Riechers et al. 2019; Decarli et al. 2020; Lenkić et al. 2020; Keating et al. 2020), with this value originally identified in three $z \sim 1.5$ galaxies (Daddi et al. 2010). Our top-line result of $\langle T\rangle < 2.4\,\mu$K corresponds to a bound of $\rho_{\rm H2} < 1.6 \times 10^8\,M_\odot\,{\rm Mpc}^{-3}$.

When we use COPSS data in combination with COMAP Season 2 data, the results change to favour higher shot noise values as shown in Fig. 3. The constraints on the clustering amplitude, whether phrased as $A_{\rm clust}$ in a b-/$\nu_{\rm eff}$-agnostic analysis or $\langle Tb\rangle^2$ in a b-/$\nu_{\rm eff}$-informed analysis, is essentially the same under COMAP Season 2 constraints with or without COPSS data. Note however that in the Early Science analyses of Chung et al. (2022), the COPSS data dominated the constraint on $P_{\rm shot}$ and weakly favoured a positive value, with the $b/\nu_{\rm eff}$-agnostic 2D probability distribution between $A_{\rm clust}$ and $P_{\rm shot}$ resembling a 2D Gaussian distribution just truncated at the 2σ contour by the $P_{\rm shot} = 0$ boundary. This is no longer the case, with the corresponding distribution in the upper portion of Fig. 3 resembling a 2D Gaussian distribution truncated inside the 1σ contour.

What greater allowance remains for higher $P_{\rm shot}$ values still comes from the way in which the b-/$\nu_{\rm eff}$-informed analysis adds an attempted correction for line broadening. Previous work by Chung et al. (2021) showed that the finite widths of line profiles can attenuate the power spectrum by $\sim 10\%$ at scales relevant to COMAP but at a higher $\sim 30\%$ level at scales surveyed by interferometric surveys like COPSS. By correcting for this attenuation, the b-/$\nu_{\rm eff}$-informed "COMAP S2+COPSS" analysis obtains an upper limit of $P_{\rm shot} < 4.8 \times 10^3\,\mu{\rm K}^2\,{\rm Mpc}^3$, which is 27% higher than the upper limit from the b-/$\nu_{\rm eff}$-agnostic "COMAP S2+COPSS" analysis of $P_{\rm shot} < 6.1 \times 10^3\,\mu{\rm K}^2\,{\rm Mpc}^3$. This difference is within the possible range of attenuation expected for the COPSS k-range given our modeling.[2]

While a combination of low clustering amplitude and high shot noise can certainly explain the current data, the b-/$\nu_{\rm eff}$-informed COMAP S2+COPSS analysis shown in Fig. 3 assigns significant probability mass within the 2σ contour to regions of parameter space with high clustering-to-shot noise ratios (particularly at low $\langle Tb\rangle^2$ values) that do not correspond to any known model. This analysis mode may thus be running into an unphysical parameter space without being grounded in a properly phrased halo model. For instance, by not marginalizing properly over possible values of $\nu_{\rm eff}$, and merely assuming a fixed average for each parameter space point, we potentially incorrectly de-bias against line broadening. We therefore move to consider the five-parameter analysis constraining $L(M_h)$, as opposed to nonspecific clustering and shot noise amplitudes.

3.2. Five-parameter analysis

Fig. 4 and Fig. 5 summarize the results of our five-parameter analysis in terms of derived quantities; we also show the posterior distributions in the original parameter space in Appendix B. First, comparing the "UM+COLDz" distribution with

[2] Curiously, however, the upward correction is similar between the COMAP S2 standalone analyses – in fact, it ends up slightly larger at 32%. This is not entirely insensible. Even small amounts of attenuation allowable within uncertainties at lower wavenumbers will correspond to a large possible range of corrections for attenuation of the shot noise component, and this lever arm from low k to high k is larger with COMAP data alone than with COPSS data added to the analysis.

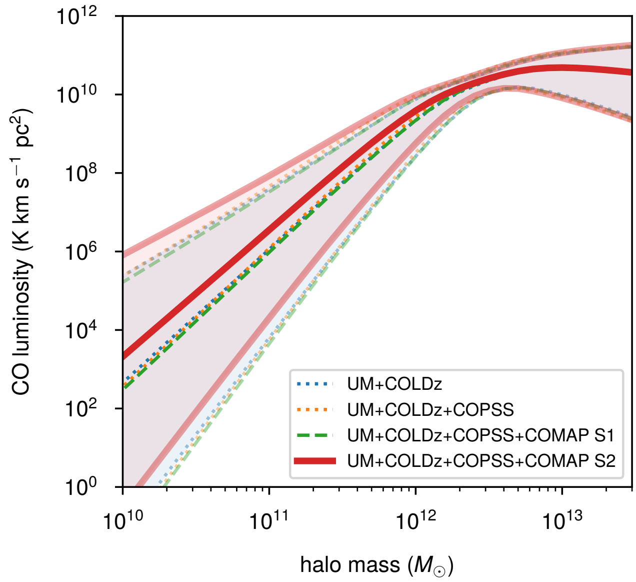


Fig. 4. 68% intervals (lighter curves) and median values (darker curves) for $L'_{\rm CO}(M_h)$ from the five-parameter MCMC described in the main text.

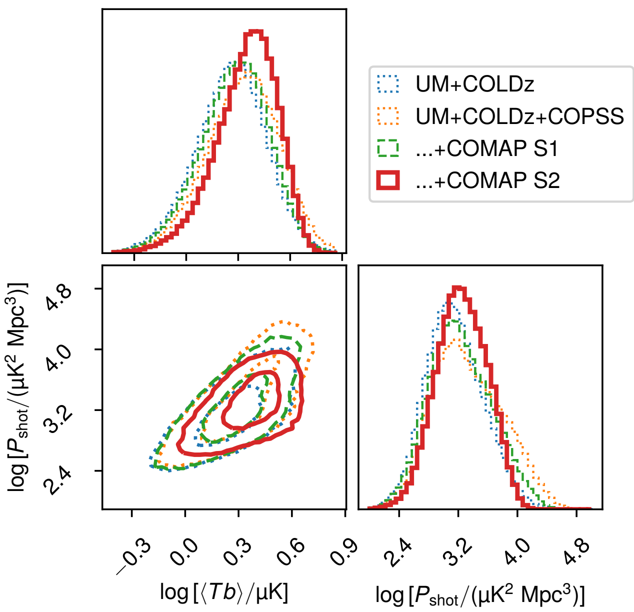


Fig. 5. Derived posterior distributions for $\langle Tb\rangle$ and $P_{\rm shot}$ based on the five-parameter MCMC described in the main text. The inner (outer) contours of each distribution show the 1σ (2σ) confidence regions.

"UM+COLDz+COPSS+COMAP S1", we do not see tangible differences in the derived quantities. The COPSS data on their own push the parameter space towards slightly higher σ, lower A (so a steeper faint-end slope for the $L(M_h)$ relation), and overall a brighter signal as shown in the posteriors for $\langle Tb\rangle$ and $P_{\rm shot}$. However, the COMAP Season 1 non-detection essentially reverses many of these changes, even suggesting a slightly dimmer faint end of the halo mass–CO luminosity relation.

The COMAP Season 2 results push expectations for the signal back up, albeit only marginally. By pushing the double power law pivot mass M to lower values and pushing A (the opposite of the faint-end slope) to higher values, the COMAP Season 2 analysis suggests a brighter population of low-mass halos than our previous data would allow, as shown in Fig. 4. This is also

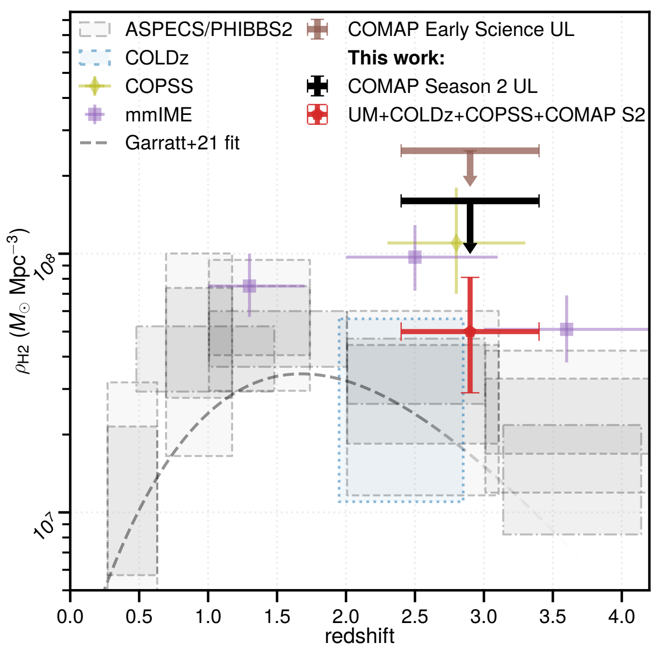


Fig. 6. Constraints on the cosmic molecular gas mass density ρ_{H2} based on CO abundance measurements across redshifts 0–4. We show COMAP Season 1 and Season 2 constraints on $\langle T \rangle$ converted based on Eq. (13), including both the two-parameter analysis upper limit and the five-parameter UM+COLDz+COPSS+COMAP S2 result. For comparison, we also show past results from untargeted interferometric CO line searches (boxes) – ASPECS (Decarli et al. 2020), PHIBBS2 (Lenkić et al. 2020), and COLDz (Riechers et al. 2019) – as well as interferometric CO LIM surveys (uncapped error bars) – COPSS (Keating et al. 2016) and mmIME (Keating et al. 2020). In addition, we show a best-fit model describing results from stacked 850 μm luminosities of galaxies at redshifts 0–2.5 (Garratt et al. 2021). All results use $\alpha_{CO} = 3.6\,M_\odot\,(\mathrm{K\,km\,s^{-1}\,pc^{-2}})^{-1}$ (cf. Daddi et al. 2010) except COPSS, which uses a conversion of $\alpha_{CO} = 4.3\,M_\odot\,(\mathrm{K\,km\,s^{-1}\,pc^{-2}})^{-1}$ (cf. Bolatto et al. 2013), and Garratt et al. (2021), who use $\alpha_{CO} = 6.5\,M_\odot\,(\mathrm{K\,km\,s^{-1}\,pc^{-2}})^{-1}$ as promoted by Scoville et al. (2016).

apparent in Fig. 5, where the "UM+COLDz+COPSS+COMAP S2" posteriors for the derived quantities $\langle Tb \rangle$, and P_{shot} show a systematic shift towards higher $\langle Tb \rangle$ – for the first time markedly pushing away from the lower limit implied by the UM+COLDz distribution – in addition to a less extended right tail for P_{shot}. These shifts in the posterior distribution suggest that the COMAP Pathfinder survey is approaching the point of making statements about the faint end of the CO luminosity function by accessing its contribution to the clustering of CO emissivity on cosmological scales, something no other survey on the horizon will do.

Finally, we note that the "UM+COLDz+COPSS+COMAP S2" estimate for $\langle T \rangle$ – which incorporates prior information and should not be considered a COMAP "detection" of any kind – is $0.72^{+0.45}_{-0.30}\,\mu$K. This corresponds to a cosmic molecular gas density of $\rho_{H2} = 5.0^{+3.1}_{-2.1} \times 10^7\,M_\odot\,\mathrm{Mpc}^{-3}$, which we discuss further in the next section.

4. Discussion

Phrased in terms of constraints on ρ_{H2}, the COMAP Season 2 results show the progress that COMAP – and thus single-dish CO LIM as a technique – has made in growing into an independent probe of cosmological CO emission and thus of cosmic molecular gas content. We illustrate this graphically in Fig. 6, showing the present work's COMAP results in the context of previous work. The results from prior literature are mostly the same as those Chung et al. (2022) collated for their Fig. 9.

- Deep surveys have leveraged community interferometers to observe pencil beam volumes and identify CO line emission candidates from the integrated data cubes in a serendipitous fashion. Of surveys used for such untargeted line searches, only the COLDz survey previously discussed in Sect. 2.3 directly observes high-redshift CO(1–0) line emission.
- Two other deep interferometric surveys – the ALMA SPECtroscopic Survey in the Hubble Ultra Deep Field (ASPECS; Decarli et al. 2020) and the Plateau de Bure High-z Blue Sequence Survey 2 (PHIBBS2; Lenkić et al. 2020) – include 3 mm observations sensitive to a range of CO lines including CO(3–2) at $z \approx 2$–3. These constraints on CO luminosity density, and thus ρ_{H2}, are subject to an additional conversion to CO(1–0) luminosity from higher-J CO lines.
- Community interferometers have also hosted key pilot small-scale CO LIM surveys, namely the previously mentioned COPSS and mmIME.

As an additional reference point, we also overplot the best-fitting model from Garratt et al. (2021) to stacked 850 μm luminosities of near-infrared selected galaxies at redshifts 0–2.5. That work took advantage of a tight empirical correlation identified by Scoville et al. (2016) between the 850 μm luminosity and CO(1–0) luminosity of both low-redshift galaxies and $z \sim 2$ submillimeter galaxies. This stands in contrast to the other results assembled, which directly survey CO lines in some fashion, although not always specifically CO(1–0).

While COMAP Season 2 data are in weak disagreement with the COPSS results, this does not translate into a disagreement in the space of ρ_{H2}. This is due to the way in which Keating et al. (2016) derived ρ_{H2} from the COPSS results. The derivation involved a number of stringent model assumptions including a linear relation between halo mass and CO luminosity, a linear relation between halo mass and molecular gas mass fraction, and the introduction of a prior on the log-normal scatter σ that suppressed the preferred amount of CO luminosity per halo mass versus what an unconstrained analysis would have found. Such assumptions motivate the analyses carried out in the present work, analysing multiple datasets through common modeling frameworks with shared assumptions. Compare, for instance, our own COPSS re-analysis which found an upper limit of $\rho_{H2} < 6.4$–$7.4 \times 10^8\,M_\odot\,\mathrm{Mpc}^{-3}$ in Sect. 3.1, versus our own COMAP S2 upper limit of $\rho_{H2} < 1.6 \times 10^8\,M_\odot\,\mathrm{Mpc}^{-3}$.

As mentioned at the end of Sect. 3.2, our best estimate for ρ_{H2} when combining COMAP Season 2 with external prior information is $\rho_{H2} = 5.0^{+3.1}_{-2.1} \times 10^7\,M_\odot\,\mathrm{Mpc}^{-3}$. We show in Fig. 6 that this "UM+COLDz+COPSS+COMAP S2" estimate lies squarely between constraints from CO line searches, which cluster lower, and constraints from interferometric CO LIM surveys, which cluster higher. These two different families of experiments informed two different sets of forecasts of COMAP Pathfinder five-year results in Chung et al. (2022), one using the fiducial data-driven "UM+COLDz+COPSS" model[3] and the other using

[3] Although this model originated from a data-driven prior that also used COPSS, the COLDz data clearly dominated the information content reflected in the prior. We see this again in the present work from the minimal difference between the "UM+COLDz" and "UM+COLDz+COPSS" posteriors in Sect. 3.2.

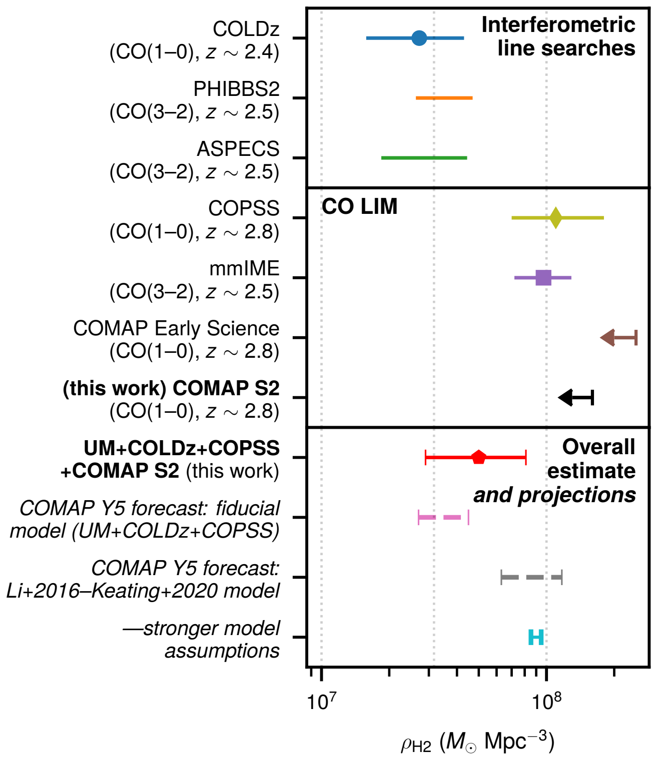


Fig. 7. Constraints on ρ_{H2} just before "cosmic noon", showing 95% upper limits and confidence intervals standardized to 68% when possible. In addition to constraints shown before in Fig. 6 from line searches (Riechers et al. 2019; Lenkić et al. 2020; Decarli et al. 2020), previous CO LIM analyses (Keating et al. 2016, 2020; Chung et al. 2022), and the present work, we also show COMAP Pathfinder five-year forecasts from Chung et al. (2022).

the Li et al. (2016)–Keating et al. (2020) model originally formulated to explain mmIME results. Our best estimate thus also lies between these two models from COMAP Early Science and the forecasts that used them, which we show alongside previous and current results in Fig. 7.

As COMAP continues to move forward as a single-dish experiment, its large-scale imaging will complement interferometric surveys in important ways. This includes not only COPSS and mmIME but also resolved line candidate searches like ASPECS, since COMAP relies on statistical large-scale fluctuations rather than individual sources. For example, ASPECS-Pilot detected 14 high-redshift [C II] line candidates (Aravena et al. 2016) only to show in the subsequent ASPECS Large Program observations that every single one was spurious (Uzgil et al. 2021). The importance of having independent single-dish LIM experiments like COMAP in the conversation will only increase as COMAP accrues further data.

The other important focus of COMAP that is salient to the wider landscape of CO abundance measurements is its focus on low-J CO lines, specifically CO(1–0) at $z \sim 3$ in the case of the Pathfinder survey. For example, comparing the results of Garratt et al. (2021) against those of ASPECS or PHIBBS2 would require accounting for not only uncertainties in quantities like α_{CO}, but also the respective conversion from the original measurement into CO(1–0) luminosity density – the Scoville et al. (2016) 850 μm–CO(1–0) conversion in the case of Garratt et al. (2021), and the conversion to CO(1–0) from higher-J CO lines observed by ASPECS and PHIBBS2 (though for ASPECS see Riechers et al. 2020). Future COMAP constraints will entirely bypass this last uncertainty by directly constraining the CO(1–0) luminosity density – and across all faint and bright galaxies in the survey volume, not constrained to any specific galaxy selection.

For now our best estimates remain consistent with all experiments, but our sensitivity to the clustering of CO has clearly improved to the point of providing upward revisions to expectations for the average CO luminosities of low-mass halos. While the current sensitivities of COMAP data to the tracer bias and average line temperature are at best marginal against our informative priors, they will continue to grow as we accrue more data. As Lunde et al. (2024) note and as we have already noted in the Introduction, the COMAP Pathfinder achieved nearly an order-of-magnitude increase in power spectrum sensitivity per k-bin despite only a 3.4× increase in raw data volume. With continued improvements in data cleaning and analysis, we remain optimistic that the amount of usable data will increase nonlinearly with further observing seasons and allow the COMAP Pathfinder to meet its targets for five-year sensitivities. As it does so, the resulting constraints will readily lend itself to very straightforward joint analyses with other measurements of cosmic CO(1–0) emissivity or molecular gas content in the vein of other LIM surveys, line scan surveys, or even analyses like that of Garratt et al. (2021), enhancing our understanding of how the rise and fall of cosmic star-forming gas relates to its depletion through the rise and fall of cosmic star-formation activity.

5. Conclusions

With the above results and discussion, we now have firm answers to the questions posed in the Introduction to this work:

- *How much does the increased data volume improve constraints on the clustering and shot noise power of cosmological CO(1–0) emission at $z \sim 3$?* The COMAP Season 2 dataset represents a five-fold improvement in upper bounds on CO shot noise power and a halving of the upper bound on the CO clustering amplitude over COMAP Early Science. This increased sensitivity introduces tension against the previous COPSS result, which will evolve with future analyses.
- *Can COMAP Season 2 results better constrain the empirical connection between CO emission and the underlying structures of dark matter?* While COMAP Season 2 data only provide marginal improvements in constraining this connection, we see hints of the COMAP Pathfinder's basic capability in capturing the clustering of low-mass CO emitters in ways that other experiments cannot.

The present work has taken an extremely conservative approach to high-level analysis, with generic models for either the power spectrum or the halo–CO connection. By making even stronger model assumptions we can make statements about semi-analytic models of galaxy formation and the connection between star-formation activity and molecular gas content (cf. Breysse et al. 2022). We leave this to a future collaboration work currently in preparation.

The outlook for the COMAP Pathfinder remains strongly positive as it continues past three years of data acquisition. The improvements demonstrated in Season 2, not only in observing efficiency but also in data cleaning and processing as demonstrated by the papers that this work accompanies (Lunde et al. 2024; Stutzer et al. 2024), will continue to grow with further Pathfinder operations. The collaboration thus continues to be on track for the outcome forecast by Chung et al. (2022): a high-significance detection of cosmological CO clustering sometime in the next few years.

Acknowledgements. DTC was supported by a CITA/Dunlap Institute postdoctoral fellowship for much of this work. The Dunlap Institute is funded through an endowment established by the David Dunlap family and the University of Toronto. The University of Toronto operates on the traditional land of the Huron-Wendat, the Seneca, and most recently, the Mississaugas of the Credit River; DTC is grateful to have the opportunity to work on this land. Research in Canada is supported by NSERC and CIFAR. This material is based upon work supported by the National Science Foundation under Grant Nos. 1517108, 1517288, 1517598, 1518282, 1910999, and 2206834, as well as by the Keck Institute for Space Studies under "The First Billion Years: A Technical Development Program for Spectral Line Observations". Parts of the work were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. PCB acknowledges support from the James Arthur Postdoctoral Fellowship during the writing of this work. We acknowledge and thank the support from the Research Council of Norway through grants 251328 and 274990, and from the European Research Council (ERC) under the Horizon 2020 Research and Innovation Program (grant agreements 772253 bits2cosmology and 819478 Cosmoglobe). HP acknowledges support from the Swiss National Science Foundation via Ambizione Grant PZ00P2_179934. This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT, No. RS-2024-00340759). This study uses the Scientific color map `acton` (Crameri 2021) to prevent visual distortion of the data and exclusion of readers with color-vision deficiencies (Crameri et al. 2020). As in Chung et al. (2022) we thank Riccardo Pavesi for access to the COLDz ABC posterior sample used in this work. Many thanks also to George Stein for running and making available the original peak-patch simulations for use with the COMAP Pathfinder survey. This research made use of NASA's Astrophysics Data System Bibliographic Services. Some of the methods for this work were originally tested on the Sherlock cluster; DTC would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. This work was first presented at the Line Intensity Mapping 2024 conference held in Urbana, Illinois; we thank Joaquin Vieira and the other organizers for their hospitality and the participants for useful discussions.

References

Aravena, M., Decarli, R., Walter, F., et al. 2016, ApJ, 833, 71
Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, ApJ, 762, L31
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013b, ApJ, 770, 57
Bernal, J. L. & Kovetz, E. D. 2022, A&A Rev., 30, 5
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207
Breysse, P. C., Yang, S., Somerville, R. S., et al. 2022, ApJ, 929, 30
Carilli, C. L. & Walter, F. 2013, ARA&A, 51, 105
Chung, D. T., Breysse, P. C., Cleary, K. A., et al. 2022, ApJ, 933, 186
Chung, D. T., Breysse, P. C., Ihle, H. T., et al. 2021, ApJ, 923, 188
Cleary, K. A., Borowska, J., Breysse, P. C., et al. 2022, ApJ, 933, 182
Crameri, F. 2021, Scientific colour maps
Crameri, F., Shephard, G. E., & Heron, P. J. 2020, Nature Communications, 11, 5444
Daddi, E., Bournaud, F., Walter, F., et al. 2010, ApJ, 713, 686
Decarli, R., Aravena, M., Boogaard, L., et al. 2020, ApJ, 902, 110
Förster Schreiber, N. M. & Wuyts, S. 2020, ARA&A, 58, 661
Foss, M. K., Ihle, H. T., Borowska, J., et al. 2022, ApJ, 933, 184
Garratt, T. K., Coppin, K. E. K., Geach, J. E., et al. 2021, ApJ, 912, 62
Hamilton, A. J. S. 1998, in Astrophysics and Space Science Library, Vol. 231, The Evolving Universe, ed. D. Hamilton, 185
Ihle, H. T., Borowska, J., Cleary, K. A., et al. 2022, ApJ, 933, 185
Kaiser, N. 1987, MNRAS, 227, 1
Kamenetzky, J., Rangwala, N., Glenn, J., Maloney, P. R., & Conley, A. 2016, ApJ, 829, 93
Keating, G. K., Marrone, D. P., Bower, G. C., & Keenan, R. P. 2020, ApJ, 901, 141
Keating, G. K., Marrone, D. P., Bower, G. C., et al. 2016, ApJ, 830, 34
Kovetz, E., Breysse, P. C., Lidz, A., et al. 2019, BAAS, 51, 101
Lamb, J. W., Cleary, K. A., Woody, D. P., et al. 2022, ApJ, 933, 183
Lenkić, L., Bolatto, A. D., Förster Schreiber, N. M., et al. 2020, AJ, 159, 190
Li, T. Y., Wechsler, R. H., Devaraj, K., & Church, S. E. 2016, ApJ, 817, 169
Lunde, J. G. S., Stutzer, N. O., Breysse, P. C., et al. 2024, arXiv e-prints, arXiv:2406.07510
Padmanabhan, H. 2018, MNRAS, 475, 1477
Pavesi, R., Sharon, C. E., Riechers, D. A., et al. 2018, ApJ, 864, 49
Pullen, A. R., Chang, T.-C., Doré, O., & Lidz, A. 2013, ApJ, 768, 15
Riechers, D. A., Boogaard, L. A., Decarli, R., et al. 2020, ApJ, 896, L21
Riechers, D. A., Pavesi, R., Sharon, C. E., et al. 2019, ApJ, 872, 7
Scoville, N., Sheth, K., Aussel, H., et al. 2016, ApJ, 820, 83
Somerville, R. S. & Davé, R. 2015, ARA&A, 53, 51
Stein, G., Alvarez, M. A., & Bond, J. R. 2019, MNRAS, 483, 2236
Stutzer, N. O., Lunde, J. G. S., Breysse, P. C., et al. 2024, arXiv e-prints, arXiv:2406.07511
Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
Uzgil, B. D., Oesch, P. A., Walter, F., et al. 2021, ApJ, 912, 67
Yang, S., Popping, G., Somerville, R. S., et al. 2022, ApJ, 929, 140

Appendix A: Alternative presentation of observed power spectra

Fig. A.1 provides an alternate presentation of the LIM observations shown in Fig. 1 as part of Sect. 2.1. Namely, we represent all $P(k)$ datasets not as upper limits given positivity priors, and instead as data points on a linear scale. In this case we take the feed or feed-group pseudo cross-power spectrum data $\tilde{C}(k)$ derived in Ihle et al. (2022) and Stutzer et al. (2024) as the best COMAP Season 1 and Season 2 estimates for the astrophysical $P(k)$. We also show credible ranges of $P(k)$ values from the two-parameter analyses of Sect. 3 rather than specific models as in Fig. 1.

While the Early Science work of Chung et al. (2022) represented COMAP Season 1 and COPSS data as co-added results across their respective k-ranges, we will not adopt such representation in future work. Inter-bin correlations, interacting with imposition of a positivity prior on the co-added result versus on the individual $P(k)$ points, could result in differences between co-added and per-bin results impossible to make sense of. For COMAP Season 2 data, we have explicitly verified that inter-bin correlations are $\lesssim 10\%$ for our chosen k-binning (Stutzer et al. 2024). Even so, the choice of how to represent a co-added result is fraught with many choices with respect to the averaging scheme, the central k-value, and can lead to misleading visual comparisons when plotting model power spectra alongside co-added results without the same weighting used to average the observed power spectra.

Appendix B: Full five-parameter MCMC posterior distributions

In Fig. B.1 we show the full five-parameter posterior distributions from the analysis of Sect. 2.3, from which we obtain distributions for the derived quantities shown in Sect. 3.2. Compared to Fig. 5, the changes due to COMAP Season 2 data are more subtle in this higher-dimensional space but nonetheless discernible.

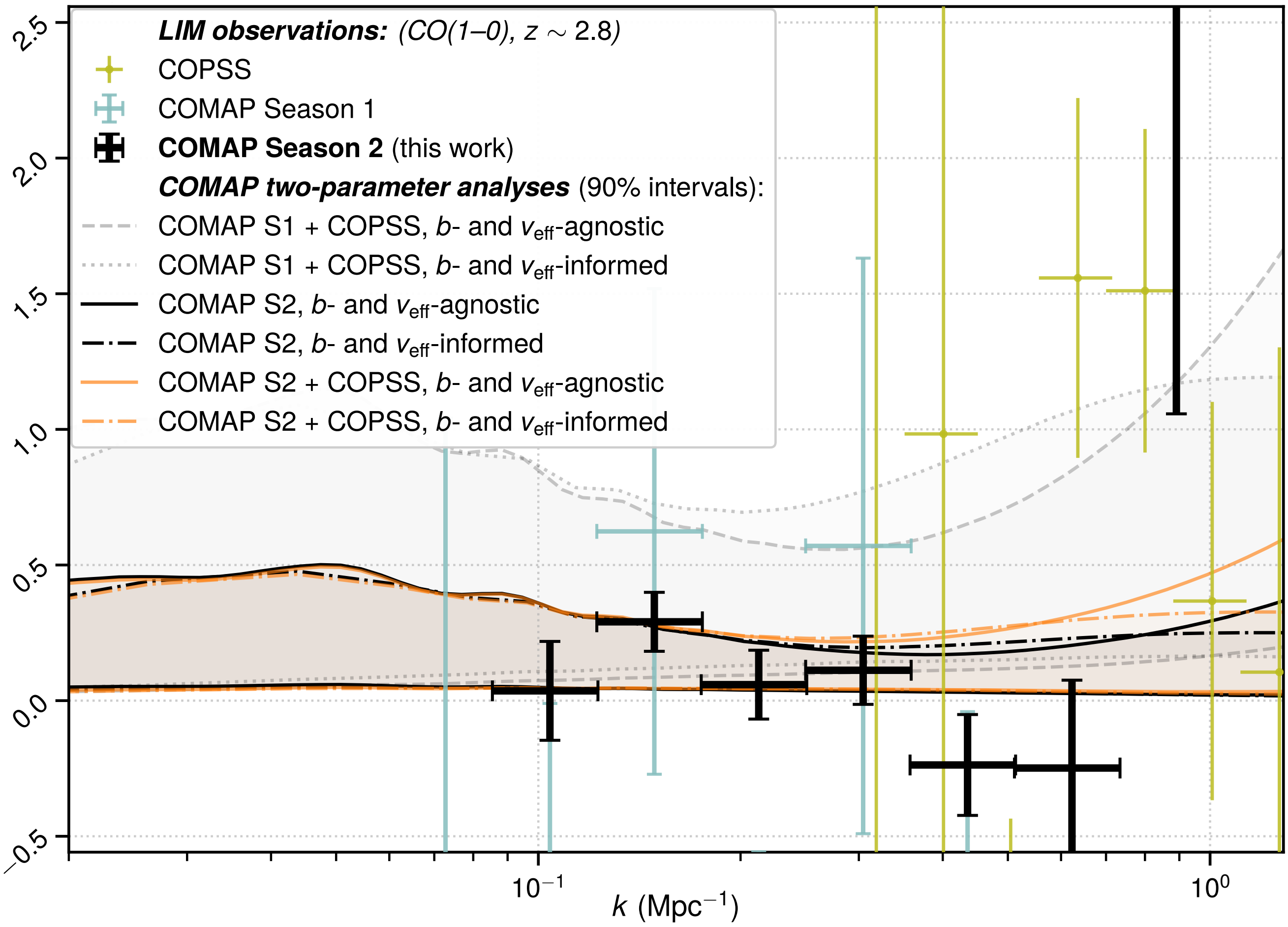


Fig. A.1. Same as Fig. 1, but with all LIM $P(k)$ results shown with 1σ uncertainties per k-bin and on a linear scale. Furthermore, instead of the specific models shown in Fig. 1, we show the typical range of allowable power spectrum values based on the two-parameter analyses of Sect. 3, with the b-/ν_{eff}-informed variations showing attenuation for line broadening left uncorrected. These allowable ranges shown should not be taken to represent a detection as they assume non-negative $P(k)$ values by definition.

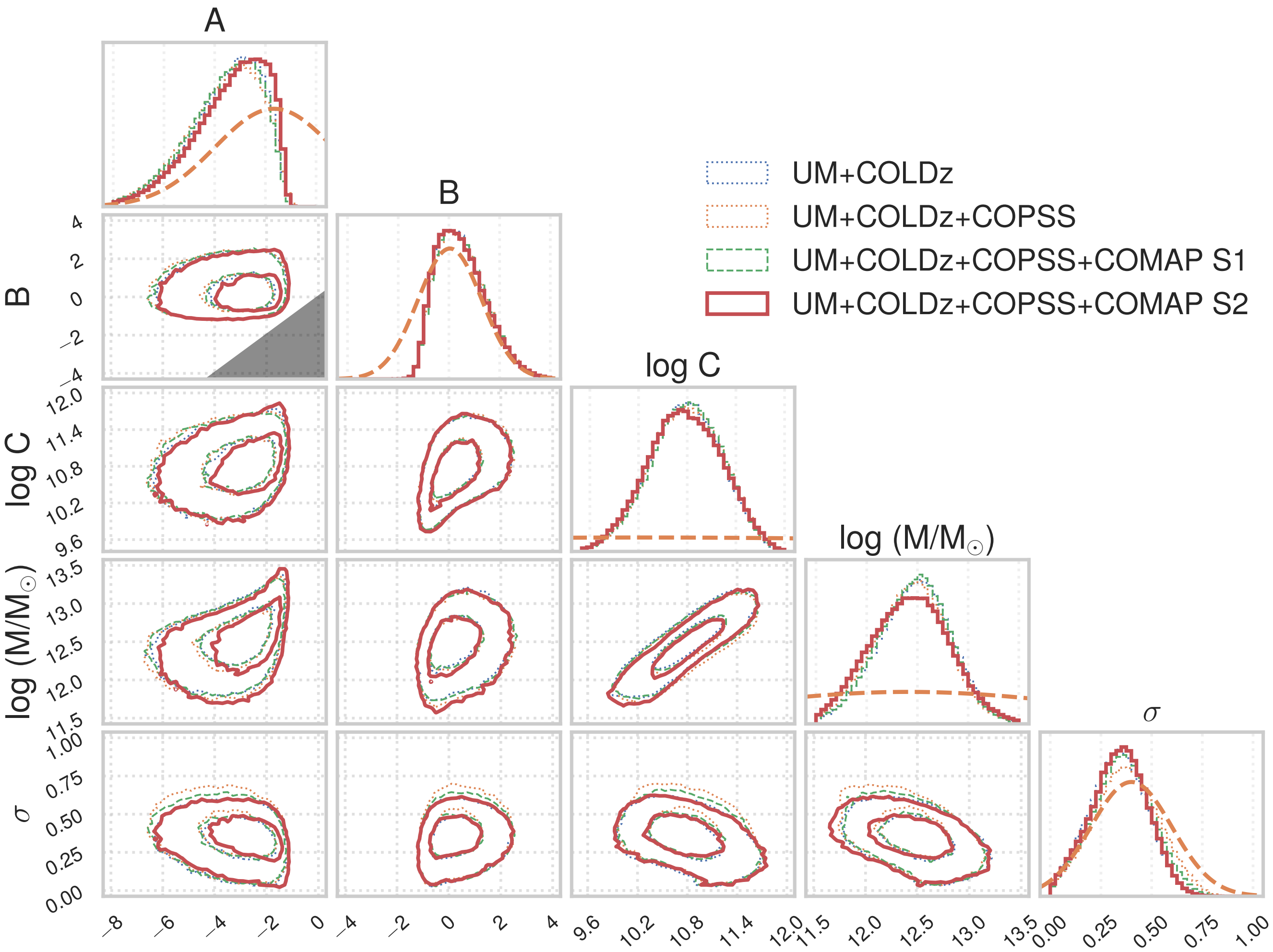


Fig. B.1. MCMC posterior distributions for the five parameters of our $L'_{\rm CO}(M_h)$ model, obtained from the analysis described in Sect. 2.3 of the main text. The inner (outer) contours of each 2D distribution show the 39% (86%) or roughly 1σ (2σ) confidence regions. The grey triangle in the 2D probability distribution between A and B shows a never-accessed region where $A > B$; the MCMC treats the two parameters as an interchangeable pair, with the smaller (larger) of the two always subjected to the prior for A (B). Dashed lines represent the loosely informative "UM" priors discussed in Sect. 2.3.

Paper VI

BeyondPlanck IV. Simulations and validation

A&A 675, A4 (2023)
https://doi.org/10.1051/0004-6361/202244958

Astronomy & Astrophysics
Special issue

BeyondPlanck: end-to-end Bayesian analysis of Planck LFI

BeyondPlanck

IV. Simulations and validation

M. Brilenkov[1,⋆], K. S. F. Fornazier[2], L. T. Hergt[3], G. A. Hoerning[2], A. Marins[2], T. Murokoshi[4], F. Rahman[5], N.-O. Stutzer[1], Y. Zhou[6], F. .B. Abdalla[2,7,8,9], K. J. Andersen[1], R. Aurlien[1], R. Banerji[1], A. Basyrov[1], A. Battista[10], M. Bersanelli[11,12,13], S. Bertocco[23], S. Bollanos[10], L. P. L. Colombo[11,13], H. K. Eriksen[1], J. R. Eskilt[1], M. K. Foss[1], C. Franceschet[11,13], U. Fuskeland[1], S. Galeotta[23], M. Galloway[1], S. Gerakakis[10], E. Gjerløw[1], B. Hensley[15], D. Herman[1], T. D. Hoang[16], M. Ieronymaki[10], H. T. Ihle[1], J. B. Jewell[17], A. Karakci[1], E. Keihänen[18,19], R. Keskitalo[20], G. Maggio[23], D. Maino[11,12,13], M. Maris[23], S. Paradiso[11], B. Partridge[21], M. Reinecke[22], A.-S. Suur-Uski[18,19], T. L. Svalheim[1], D. Tavagnacco[23,14], H. Thommesen[1], M. Tomasi[11,12], D. J. Watts[1], I. K. Wehus[1], and A. Zacchei[23]

(Affiliations can be found after the references)

Received 12 September 2022 / Accepted 2 April 2023

ABSTRACT

End-to-end simulations play a key role in the analysis of any high-sensitivity cosmic microwave background (CMB) experiment, providing high-fidelity systematic error propagation capabilities that are unmatched by any other means. In this paper, we address an important issue regarding such simulations, namely, how to define the inputs in terms of sky model and instrument parameters. These may either be taken as a constrained realization derived from the data or as a random realization independent from the data. We refer to these as posterior and prior simulations, respectively. We show that the two options lead to significantly different correlation structures, as prior simulations (contrary to posterior simulations) effectively include cosmic variance, but they exclude realization-specific correlations from non-linear degeneracies. Consequently, they quantify fundamentally different types of uncertainties. We argue that as a result, they also have different and complementary scientific uses, even if this dichotomy is not absolute. In particular, posterior simulations are in general more convenient for parameter estimation studies, while prior simulations are generally more convenient for model testing. Before BeyondPlanck, most pipelines used a mix of constrained and random inputs and applied the same hybrid simulations for all applications, even though the statistical justification for this is not always evident. BeyondPlanck represents the first end-to-end CMB simulation framework that is able to generate both types of simulations and these new capabilities have brought this topic to the forefront. The BeyondPlanck posterior simulations and their uses are described extensively in a suite of companion papers. In this work, we consider one important applications of the corresponding prior simulations, namely, code validation. Specifically, we generated a set of one-year LFI 30 GHz prior simulations with known inputs and we used these to validate the core low-level BeyondPlanck algorithms dealing with gain estimation, correlated noise estimation, and mapmaking.

Key words. cosmic background radiation – cosmology: observations – diffuse radiation

1. Introduction

High-fidelity end-to-end simulations play a critical role in the analysis of any modern cosmic microwave background (CMB) experiment for at least three important reasons. Firstly, during the design phase of the experiment, simulations are used to optimize and forecast the performance of a given experimental design and ensure that the future experiment will achieve its scientific goals (e.g., LiteBIRD Collaboration 2023). Secondly, simulations are essential for validation purposes, since they may be used to test data-processing techniques as applied to a realistic instrument model. Thirdly, realistic end-to-end simulations play an important role in bias and error estimations for traditional CMB analysis pipelines.

Simulations played a particularly important role in the data reduction of *Planck* and massive efforts were invested in implementing efficient and re-usable analysis codes that were generally applicable to a wide range of experiments. This work started with the LevelS software package (Reinecke et al. 2015) and culminated with the Time Ordered Astrophysics Scalable Tools[1] (TOAST), which was explicitly designed to operate in a massively parallel high-performance computing environment. TOAST was used to produce the final generations of the *Planck* Full Focal Plane (FFP) simulations (Planck Collaboration XII 2016), which served as the main error propagation mechanism in the *Planck* 2015 and 2018 data releases (Planck Collaboration I 2016, 2020).

For *Planck*, generating end-to-end simulations have represented (by far) the dominant computational cost of the entire experiment, accounting for 25 million CPU-hrs in the 2015 data release alone. In addition, the production phase required massive amounts of human effort, in terms of preparing the inputs, executing the runs, and validating the outputs. It is of great interest for any future experiment to optimize and streamline this simulation process, and reuse both validated software and human work whenever possible.

In this respect, the BeyondPlanck end-to-end Bayesian analysis framework (BeyondPlanck Collaboration 2023) offers a novel approach to generating CMB simulations. While the

⋆ Corresponding author: M. Brilenkov,
e-mail: maksym.brilenkov@astro.uio.no

[1] https://github.com/hpc4cmb/toast

primary goal of this framework is to draw samples from a full joint posterior distribution for analytical purposes, it is useful to note that the foundation of this approach is simply a general and explicit parametric model for the full time-ordered data (TOD). When exploring the full joint posterior distribution, this model is compared with the observed data in the TOD space. The analysis phase is as such numerically equivalent to producing a large number of TOD simulations and comparing each of these with the actual observed data. In this framework, each step of the analysis and simulation pipelines are thus fully equivalent, and the primary difference is simply based on whether the input model parameters are assumed to be constrained by the data or not.

This latter observation is in fact a key point regarding end-to-end simulations for CMB experiments in general, and a main goal of the current paper is to clarify the importance of choosing input parameters for a given simulation appropriately. Specifically, we argue in this paper that two fundamentally different choices are available: we can either choose parameters that are constrained directly by the observed data (as is traditionally done for the CMB Solar dipole or astrophysical foregrounds) or we can choose parameters that are independent from the observed data (as is traditionally done for CMB fluctuations or instrumental noise). We further argue that this choice will have direct consequences for the specific scientific questions the resulting simulations are optimized to address.

It is important to note that these ideas were discussed broadly, but not systematically, within the *Planck* community before building the FFP simulations. For instance, one proposal was to base the large-scale CMB temperature fluctuations at $\ell \leq 70$ from constrained WMAP realizations (Bennett et al. 2013), and thereby integrate knowledge about the real sky into the simulations. Another proposal was to use the actually observed LFI gain measurements to generate the simulations. A third and long-standing discussion has revolved around which values to adopt for the CMB Solar dipole.

The BeyondPlanck framework offers a novel systematic view on these questions, as our Bayesian approach provides for the first time statistically well-defined constrained realizations for all parameters in the sky model – and not just a small subset. Furthermore, when comparing the correlation structures that arise from the posterior samples with those derived from traditional simulations, obvious and important differences appear, both in terms of the frequency maps (Basyrov et al. 2023) and CMB maps (Colombo et al. 2023).

The first main goal of the current paper is to explain these differences intuitively and in that process, we introduce the concepts of "posterior simulations" and "prior simulations"[2]. Posterior simulations are random samples drawn from $P(\omega|\boldsymbol{d})$, and represent simulations that are constrained by the observed data; these are thus identical to the posterior samples described by BeyondPlanck Collaboration (2023). In contrast, prior simulations drawn from a prior distribution, $P(\omega)$, and are, as such, unconstrained by the data. We note that a similar distinction has recently been made in terms of a so-called "Bayesian workflow" by Betancourt[3] and Gelman et al. (2020).

The second main goal of this paper is simply to demonstrate in practice how the BeyondPlanck machinery may be used to generate prior simulations, on a similar footing as TOAST, and we will use these simulations for one important application, namely code validation. As discussed by Galloway et al. (2023a) and Gerakakis et al. (2023), the `Commander` code that forms the computational basis of the BeyondPlanck pipeline is explicitly designed to be re-used for a wide range of experiments. It is therefore critically important that this implementation is thoroughly validated with respect to statistical bias and uncertainties. We do so by analyzing well-controlled simulations in this paper.

At the same time, we note that the use of prior simulations is not a requirement for the validation study, as any one of the posterior simulations would have served equally well as an input for the TOD generation process. Rather, our main motivation for using prior simulations for this particular task is simply that the posterior simulations have already been used extensively in many companion papers.

The rest of the paper is organized as follows. We first provide a brief overview of the BeyondPlanck framework and data model in Sect. 2. In Sect. 3, we introduce the concept of posterior and prior simulations, and we discuss their difference. In Sect. 4, we describe the input parameters and simulation configuration used in this paper, before using these simulations to validate the BeyondPlanck implementation in Sect. 5. We present our conclusions in Sect. 6.

2. BeyondPlanck data model and Gibbs sampler

As described in BeyondPlanck Collaboration (2023) and its companion papers, the single most fundamental component of the BeyondPlanck framework is an explicit parametric model that is to be fitted to raw TOD that includes instrumental, astrophysical, and cosmological parameters. For the current analysis, this model takes the following form:

$$d_{j,t} = g_{j,t}\mathsf{P}_{tp,j}\left[\mathsf{B}^{\mathrm{symm}}_{pp',j}\sum_{c}\mathsf{M}_{cj}(\beta_{p'},\Delta_{\mathrm{bp}}{}^{j})a^{c}_{p'} + \mathsf{B}^{4\pi}_{j,t}\boldsymbol{s}^{\mathrm{orb}}_{j} + \mathsf{B}^{\mathrm{asymm}}_{j,t}\boldsymbol{s}^{\mathrm{fsl}}_{t}\right] \quad (1)$$
$$+ a_{1\,\mathrm{Hz}}\boldsymbol{s}^{1\,\mathrm{Hz}}_{j} + n^{\mathrm{corr}}_{j,t} + n^{\mathrm{w}}_{j,t},$$

where p denotes a single pixel on the sky, and c represents one single astrophysical signal component. Furthermore, $d_{j,t}$ denotes the measured data; $g_{j,t}$ denotes the instrumental gain; $\mathsf{P}_{tp,j}$ is a pointing matrix; $\mathsf{B}_{pp',j}$ denotes beam convolution with either the (symmetric) main beam, the (asymmetric) far sidelobes, or the full 4π beam response; $\mathsf{M}_{cj}(\beta_p, \Delta_{\mathrm{bp}})$ denotes the so-called mixing matrix, which describes the amplitude of component c as seen by radiometer j relative to some reference frequency when assuming some set of bandpass correction parameters Δ_{bp}; a^c_p is the amplitude of component c in pixel p; $s^{\mathrm{orb}}_{j,t}$ is the orbital CMB dipole signal, including relativistic quadrupole corrections; $s^{\mathrm{fsl}}_{j,t}$ denotes the contribution from far sidelobes; $s^{1\,\mathrm{Hz}}_{j,t}$ denotes the contribution from electronic 1 Hz spikes; $n^{\mathrm{corr}}_{j,t}$ denotes correlated instrumental noise; and $n^{\mathrm{w}}_{j,t}$ is uncorrelated (white) instrumental noise. The sky model, denoted by the sum over components, c, in the above expression may be written out as an explicit sum over CMB, synchrotron, free-free, AME, thermal dust, and point source emission, as described by Andersen et al. (2023), Svalheim et al. (2023b).

On the instrumental side, the correlated noise is associated with a covariance matrix, $\mathsf{N}^{\mathrm{corr}} = \langle \boldsymbol{n}^{\mathrm{corr}}(\boldsymbol{n}^{\mathrm{corr}})^T\rangle$, which may be approximated as piecewise stationary, and with a Fourier space power spectral density (PSD), $\mathsf{N}_{ff'} = P(f)\delta_{ff'}$, that for BeyondPlanck consists of a sum of a classic $1/f$ term and a

[2] Other possible names could be "Bayesian" and "frequentist" simulations.
[3] `https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html`

log-normal term (Ihle et al. 2023),

$$P(f)=\sigma_0^2\left[1+\left(\frac{f}{f_{\mathrm{knee}}}\right)^{\alpha}\right]+A_{\mathrm{p}}\exp\left[-\frac{1}{2}\left(\frac{\log_{10}f-\log_{10}f_{\mathrm{p}}}{\sigma_{\mathrm{dex}}}\right)^2\right]. \tag{2}$$

We define $\xi_n=\{\sigma_0,\alpha,f_{\mathrm{knee}},A_p\}$ as a composite parameter that is internally sampled iteratively through an individual Gibbs step, as described by Ihle et al. (2023); the peak location and width parameters of the log-normal term, f_{p} and σ_{dex}, are currently fixed at representative values.

Denoting the set of all free parameters in Eqs. (1)–(2) by ω, we can simplify Eq. (1) symbolically to

$$d_{j,t}=s^{\mathrm{tot}}_{j,t}(\omega)+n^{\mathrm{w}}_{j,t}. \tag{3}$$

The BeyondPlanck approach to CMB analysis simply amounts to mapping out the posterior distribution as given by Bayes' theorem,

$$P(\omega\mid\boldsymbol{d})=\frac{P(\boldsymbol{d}\mid\omega)P(\omega)}{P(\boldsymbol{d})}\propto\mathcal{L}(\omega)P(\omega), \tag{4}$$

where $P(\boldsymbol{d}\mid\omega)\equiv\mathcal{L}(\omega)$ is called the likelihood, $P(\omega)$ is some set of priors, and $P(\boldsymbol{d})$, the so-called evidence, is effectively a normalization constant for purposes of evaluating ω. The likelihood is easily defined, and given by Eq. (3) under the assumption that $\boldsymbol{n}^{\mathrm{w}}_j$ is Gaussian distributed,

$$-2\ln\mathcal{L}(\omega)=\left(\boldsymbol{d}-\boldsymbol{s}^{\mathrm{tot}}(\omega)\right)^t\mathsf{N}_{\mathrm{wn}}^{-1}\left(\boldsymbol{d}-\boldsymbol{s}^{\mathrm{tot}}(\omega)\right). \tag{5}$$

The prior is not as well defined, and we adopt in practice a combination of informative and algorithmic priors in the BeyondPlanck analysis (see BeyondPlanck Collaboration (2023) for an overview).

To explore this distribution by Markov chain Monte Carlo, we use the following Gibbs sampling chain (BeyondPlanck Collaboration 2023),

$$\boldsymbol{g}\leftarrow P(\boldsymbol{g}\mid\boldsymbol{d},\ \xi_n,\ \boldsymbol{a}^{1\mathrm{Hz}},\ \Delta_{\mathrm{bp}},\ \boldsymbol{a},\beta,\ C_\ell), \tag{6}$$
$$\boldsymbol{n}_{\mathrm{corr}}\leftarrow P(\boldsymbol{n}_{\mathrm{corr}}\mid\boldsymbol{d},\ \boldsymbol{g},\ \xi_n,\ \boldsymbol{a}^{1\mathrm{Hz}},\ \Delta_{\mathrm{bp}},\ \boldsymbol{a},\ \beta,\ C_\ell), \tag{7}$$
$$\xi_n\leftarrow P(\xi_n\mid\boldsymbol{d},\ \boldsymbol{g},\ \boldsymbol{n}_{\mathrm{corr}},\ \boldsymbol{a}^{1\mathrm{Hz}},\ \Delta_{\mathrm{bp}},\ \boldsymbol{a},\ \beta,\ C_\ell), \tag{8}$$
$$\boldsymbol{a}^{1\mathrm{Hz}}\leftarrow P(\boldsymbol{a}^{1\mathrm{Hz}}\mid\boldsymbol{d},\ \boldsymbol{g},\ \boldsymbol{n}_{\mathrm{corr}},\ \xi_n,\ \Delta_{\mathrm{bp}},\ \boldsymbol{a},\ \beta,\ C_\ell), \tag{9}$$
$$\Delta_{\mathrm{bp}}\leftarrow P(\Delta_{\mathrm{bp}}\mid\boldsymbol{d},\ \boldsymbol{g},\ \boldsymbol{n}_{\mathrm{corr}},\ \xi_n,\ \boldsymbol{a}^{1\mathrm{Hz}},\ \boldsymbol{a},\ \beta,\ C_\ell), \tag{10}$$
$$\beta\leftarrow P(\beta\mid\boldsymbol{d},\ \boldsymbol{g},\ \boldsymbol{n}_{\mathrm{corr}},\ \xi_n,\ \boldsymbol{a}^{1\mathrm{Hz}},\ \Delta_{\mathrm{bp}},\ C_\ell), \tag{11}$$
$$\boldsymbol{a}\leftarrow P(\boldsymbol{a}\mid\boldsymbol{d},\ \boldsymbol{g},\ \boldsymbol{n}_{\mathrm{corr}},\ \xi_n,\ \boldsymbol{a}^{1\mathrm{Hz}},\ \Delta_{\mathrm{bp}},\ \beta,\ C_\ell), \tag{12}$$
$$C_\ell\leftarrow P(C_\ell\mid\boldsymbol{d},\ \boldsymbol{g},\ \boldsymbol{n}_{\mathrm{corr}},\ \xi_n,\ \boldsymbol{a}^{1\mathrm{Hz}},\ \Delta_{\mathrm{bp}},\ \boldsymbol{a},\ \beta\ ), \tag{13}$$

where the symbol $\leftarrow$ denotes setting the variable on the left-hand side equal to a sample from the distribution on the right-hand side. In these expressions, it is worth noting that not all conditioned parameters are explicitly used in each sampling steps. For instance, the CMB power spectrum only depends conditionally on the CMB map and, therefore, $P(C_\ell\mid\boldsymbol{d},\boldsymbol{g},\boldsymbol{n}_{\mathrm{corr}},\xi_n,\boldsymbol{a}^{1\,\mathrm{Hz}},\Delta_{\mathrm{bp}},\boldsymbol{a},\beta)=P(C_\ell|\boldsymbol{a}^{\mathrm{CMB}})$, as, discussed by Wandelt et al. (2004), Eriksen et al. (2004). However, $\boldsymbol{a}^{\mathrm{CMB}}$ depends on many of the additional variables and the above full notation makes the "correlations-through-conditionals" Gibbs sampling nature of the algorithm explicit.

3. Posterior versus prior simulations

End-to-end TOD simulations have become the de facto industry standard for producing robust error estimates for high-precision experiments (e.g., Planck Collaboration XII 2016), and the data model defined in Eqs. (1)–(2) represents a succinct simulation recipe for producing such simulations: If ω is assumed to be perfectly known, then these equations can be evaluated in a forward manner without the need for parameter estimation or inversion algorithms. Then, the only stochastic terms are the correlated and white noise, both of which can be easily generated by a combination of standard random Gaussian number generators and Fourier transforms.

However, in practice ω is of course not perfectly known and the matter of precisely how ω is specified has direct and strong implications regarding what kind of information the resulting simulations can offer the user; for an example of this within the context of *Planck* LFI, we refer to Basyrov et al. (2023). In short, the key discriminator is whether ω is defined using real observed data (and, in practice, drawn from the posterior distribution, $P(\omega\mid\boldsymbol{d})$) or whether it is drawn from a data-independent hyper-distribution, for instance: informed by theoretical models and/or ground-based laboratory measurements. We will refer to these two approaches as "posterior-" and "prior-based," respectively, indicating whether or not they are conditioned on the true data in question.

We note that both posterior and prior simulations specifically refer to time-ordered data in the current paper – and not to pixelized maps or higher-level products. That is to say, we distinguish between simulation pipelines, which transform ω into timelines, and analysis pipelines, which transform timelines into higher ordered products, such as maps and power spectra.

3.1. Bayesian versus frequentist statistics

Before comparing the two simulation types through a few worked examples, it is useful to recall the fundamental difference between Bayesian and frequentist statistics, which may be summarized as follows: In frequentist statistics, the model, $\mathcal{M}$, and its parameters, ω, are considered to be fixed and known, while the data, $\boldsymbol{d}$, are considered to be the main uncertain quantity. In Bayesian statistics, on the other hand, $\boldsymbol{d}$ is assumed to be perfectly known and essentially defined by a list of numbers recorded by a measuring device, while ω is assumed to be the main unknown quantity.

This difference has important consequences for how each framework typically approaches statistical inference, and which questions they are most suited to answer. This is perhaps most easily illustrated through their most typical mode of operations. First, the classical frequentist approach to statistical inference is to construct an ensemble of simulated data sets, $\boldsymbol{d}_i$, each with parameters drawn independently from $\mathcal{M}(\omega)$. The next step is to define some statistic, $\gamma(\boldsymbol{d}_i):\mathbb{R}^N\rightarrow\mathbb{R}$, that isolates and highlights the important piece of information that the user is interested in; widely used CMB examples include χ^2 statistics, angular power spectrum statistics, or non-Gaussianity statistics. Finally, we go on to compute γ both for the simulations and the actual data and determine the relative frequency for which $\gamma(\boldsymbol{d}_{\mathrm{real}})<\gamma(\boldsymbol{d}_i)$, which is often called the p-value or "probability-to-exceed" (PTE). Values between about 0.025 and 0.975 are taken to suggest that the data are consistent with the model, while more extreme values indicate a discrepancy.

Given this prescription, it is clear that the frequentist approach is particularly suited for model testing applications; it intrinsically and directly addresses the question of whether the data are consistent with the model. As such it has been widely used in the CMB field for instance for studies of non-Gaussianity and isotropy. In this case, the null-hypothesis is easy to specify, namely that the universe is isotropic and homogeneous, and filled with Gaussian random fluctuations drawn from a ΛCDM universe with given parameters. Agreement between this null-hypothesis and the real observations is then typically assessed by computing the p-value of some preferred statistic.

Establishing some statistic that shows that the observed data are inconsistent with this hypothesis would constitute evidence of new physics and is, as such, a high-priority scientific target. In contrast, Bayesian statistics takes a fundamentally different approach to statistical inference. In this case, we consider ω to be a stochastic and unknown quantity and we want to understand how the observed data constrains ω. The most succinct summary of this is the posterior probability distribution itself, $P(\omega \mid \boldsymbol{d})$, and the starting point for this framework is therefore Bayes' theorem, as given in Eq. (4). Thus, the majority of applications of modern Bayesian statistics simply amounts to mapping out $P(\omega \mid \boldsymbol{d})$ as a function of ω by any means necessary.

At the same time, it is important to note that the likelihood $\mathcal{L}(\omega) = P(\boldsymbol{d} \mid \omega)$ on the right-hand side of Eq. (4) is a fully classical frequentist statistic, in which ω is assumed to be perfectly known and the data are uncertain. Still, it is important to note that the free parameter in $\mathcal{L}(\omega)$ is indeed ω, not $\boldsymbol{d}$, and $\mathcal{L}$ itself is really just a frequentist statistic that measures the overall goodness-of-fit between the data and the model. This statistic may then be used to estimate ω within a strictly frequentist framework. One popular example of this within the CMB field is the so-called profile likelihood.

Likewise, the Bayesian approach is also able to address the model selection problem, and this is typically done using the evidence factor, $P(\boldsymbol{d})$, in Eq. (4). The importance of this factor becomes obvious when explicitly acknowledging that all involved probability distributions in Eq. (4) actually depend on the overall model $\mathcal{M}$, and not only the individual parameter values:

$$P(\omega \mid \boldsymbol{d}, \mathcal{M}) = \frac{P(\boldsymbol{d} \mid \omega, \mathcal{M}) P(\omega \mid \mathcal{M})}{P(\boldsymbol{d} \mid \mathcal{M})}. \tag{14}$$

Mathematically, $P(\boldsymbol{d} \mid \mathcal{M})$ is simply given by the average likelihood integrated over all allowed parameter values, and classical Bayesian model selection between models $\mathcal{M}_1$ and $\mathcal{M}_2$ proceeds simply by evaluating $P(\boldsymbol{d} \mid \mathcal{M}_1)/P(\boldsymbol{d} \mid \mathcal{M}_2)$; the model with the higher evidence is preferred.

In summary, the foundational assumptions underlying frequentist and Bayesian methods are different and complementary, and they fundamentally address different questions. Frequentist statistics are ideally suited to address model testing problems (e.g., "is the observed CMB sky Gaussian and isotropic?"), while Bayesian statistics are ideally suited to address parameter estimation problems (e.g., what the best-fit ΛCDM parameters would be). At the same time, this dichotomy is by no means absolute and either framework is fully capable of addressing both types of questions if they are carefully addressed.

3.2. Constrained versus random input parameters in CMB simulations

We now return to the issue raised in the introduction to this section, namely how to properly choose ω for CMB inference based on end-to-end simulations. As discussed by Basyrov et al. (2023), essentially all CMB analysis pipelines prior to BeyondPlanck have adopted a mixture of data-constrained and data-independent parameters for this purpose. Key examples of the former are the CMB Solar dipole and Galactic foregrounds, both of which are strongly informed by real measurements. Correspondingly, classical examples of the latter are CMB fluctuations, which are typically drawn as Gaussian realizations from a ΛCDM power spectrum, and instrumental noise, which is often based on laboratory measurements. In our notation, these simulations qualify thus neither as pure posterior-based nor pure prior-based, but rather as a mixture of the two.

In contrast, each sample of ω produced by the BeyondPlanck Gibbs chain summarized in Eqs. (6)–(13) represents one possible simulated realization in which all sub-parameters in ω are determined exclusively by the real posterior distribution. This not only refers to the CMB dipole and Galactic model, but also those parameters that are traditionally chosen from external sources in classical pipelines, such as the CMB anisotropies and the specific noise realization.

The difference between these two types of simulation inputs is illustrated in Fig. 1, which compares ten independent prior time-domain realizations (red curves) with ten independent posterior realizations (black curves). The top and bottom panels show the correlated noise n_{corr} and the sky model s_{sky}, respectively, both plotted as a function of time. Starting with the frequentist simulations, we see that these are entirely uncorrelated between realizations and scatter randomly with some model-specific mean and variance. In particular, the frequentist simulations include so-called cosmic variance, that is, independent realizations have different CMB and noise amplitudes and phases, even if they are drawn from the same underlying stochastic model. In contrast, posterior simulations do not include cosmic variance, but rather focus exclusively on structures in the real data. For the sky signal component shown in the top panel of Fig. 1, this is seen in terms of two different aspects. First, the structure of all ten realizations follow very closely the same overall structure, and this is defined by the specific CMB pattern of the real sky. However, they also explicitly account for the uncertainty in the sky value at each pixel, and this is seen by the varying width of the black band. In the middle of the plot, the width is small, and this implies that the sky has been aptly measured here (due to deep scanning), while along the edges of the plot the width is larger and this implies that the sky has not been aptly measured. The variation between posterior simulations thus directly quantify the uncertainty of the true data. Intuitively speaking, this point may be summarized as follows: Uncertainties measured by frequentist simulations quantify the expected variations as observed with a random instrument in a random universe, while posterior simulations quantify the expected variations of the real instrument in the real universe.

These intuitive differences translate directly into both qualitatively and quantitatively different ensemble properties for the resulting simulations, and correspondingly also into different resulting error estimates. As a real-world illustration of this, Fig. 2 shows slices through the empirical low-resolution polarization covariance matrix computed for each of the three *Planck* LFI frequency channels using three different generations of LFI simulations, namely (from left to right columns): *Planck* 2018 (Planck Collaboration II 2020), *Planck* PR4 (Planck Collaboration Int. LVII 2020), and BeyondPlanck (BeyondPlanck Collaboration 2023). Row sections show results for the 30, 44, and 70 GHz channels, respectively, and within each section the two rows show the QQ and UQ segments

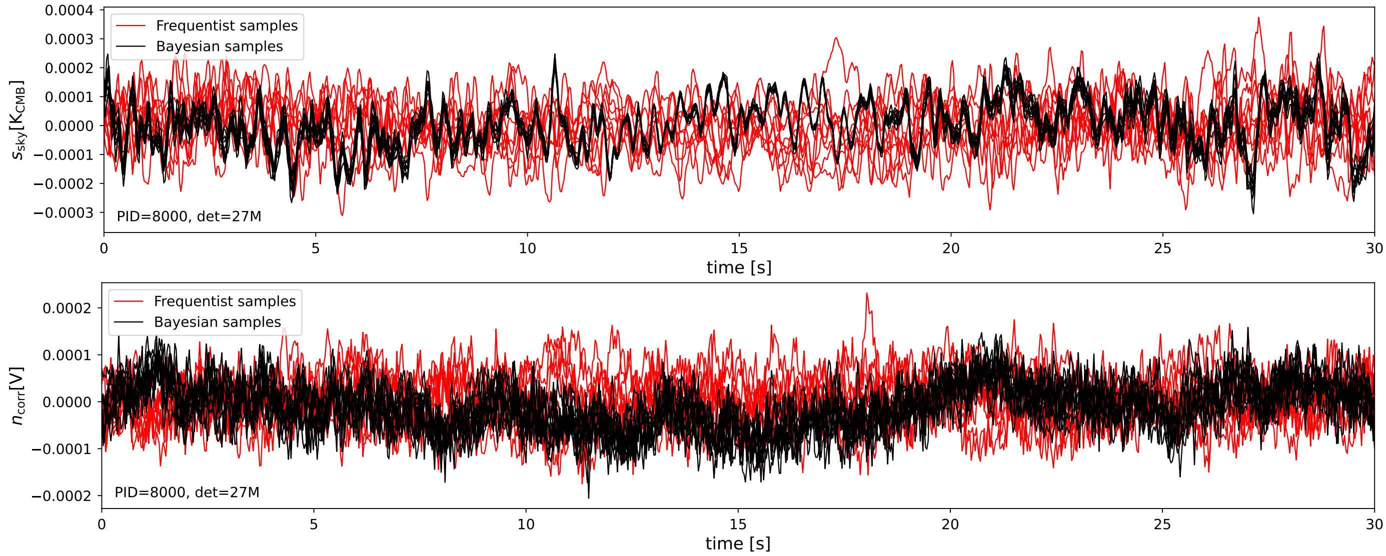


Fig. 1. Comparison of ten prior (red) and ten posterior (black) simulations in time-domain. Each line represents one independent realization of the respective type. The top panel shows sky model (i.e., CMB) simulations and the bottom panel shows correlated noise simulations.

of the full matrix, sliced through Stokes Q pixel number 100, marked in gray in the upper right quadrant. Each covariance matrix is computed by first downgrading each simulation to a HEALPix[4] (Górski et al. 2005) resolution of $N_{\rm side} = 8$, and averaging the outer product over all available realizations; we refer to Basyrov et al. (2023), Colombo et al. (2023) for further details. Effectively, these matrices visually summarize the map-space uncertainty estimates predicted by each simulation set.

Starting with the *Planck* 2018 simulations, the most striking observation is that these empirical matrices are very noisy for all three frequency channels. This is partly a reflection of the fact that only 300 simulations were actually constructed, and this leads to a high Monte Carlo uncertainty. Specifically, the uncertainty due to a finite number of simulations scales as $1/\sqrt{N_{\rm sim}}$, which suggests a 6% contribution for 300 simulations. However, because these simulations are prior-based, that number applies to all sources of variations between realizations, including white noise, instrumental effects, and sky-signal variations. Furthermore, the gains that were assumed when generating these simulations exhibited significantly less structure than the real observations. In summary, there are relatively little common structures between the various realizations, either from the astrophysical sky, the instrumental noise, or the gain, and the corresponding covariance structures are therefore weak. Visually speaking, perhaps the most notable feature is a positive correlation from correlated noise along the scanning direction that passes through the sliced pixel seen in the upper right quadrant, but these are significantly obscured by Monte Carlo uncertainties.

Proceeding to the *Planck* PR4 simulations summarized in the middle column, we now see very strong coherent structures for the 30 GHz channel, while the 44 and 70 GHz channels behave similarly to the 2018 case. The explanation for this qualitative difference is the *Planck* PR4 calibration algorithm; in this pipeline, the 30 GHz channel is calibrated independently without the use of supporting priors, while the 44 and 70 GHz channels are calibrated by using the 30 GHz channel as a polarized foreground prior. The net effect of this independent calibration procedure is a very high calibration uncertainty for the 30 GHz channel, and these couple directly to the true CMB dipole, which is kept fixed between all simulations. The result is the familiar large-scale pattern seen in this figure, which has been highlighted by several previous analyses as a particularly difficult mode to observe with *Planck* (e.g., Planck Collaboration II 2020; Gjerløw et al. 2023; Watts et al. 2023).

Turning to the BeyondPlanck simulations summarized in the right column, we now see coherent and signal-dominated structures across the full sky in all frequency channels. A part of this is simply due to more realizations than for the other two pipelines (in this case, 3200), but even more importantly, the simulations are now entirely data-driven. That is, they correspond to the black curves in Fig. 1, while the previous pipelines correspond to the red curves. In practice, this has two main effects. First, it implies that the total parameter volume that needs to be explored by Monte Carlo sampling is intrinsically smaller, simply because the posterior distribution does not include cosmic variance; the simulations only need to describe our instrument and universe – and not simply any instrument and universe, and this is a much smaller sub-set. Second, and even more importantly, the posterior simulations account naturally for non-linearity between the various parameters, and these are very often the dominant contributions in these distributions. As a concrete example, if the gain happens to scatter either high or low during a given time period, then the total uncertainty estimate will be particularly sensitive to the CMB dipole during the same time period, and it will excite a correlation structure in these plots that is intimately connected to the satellite scanning strategy. Thus, if one chooses a gain profile that is independent of other parameters, then those real uncertainties will not be properly accounted for in the simulation set: intuitively speaking, the hot and cold spots in the covariance matrices shown in Fig. 2 will either appear in the wrong places or be suppressed when averaging over independent realizations. In general, specifying the instrumental model at a sufficiently realistic level represents a real challenge for frequentist simulations, and great care is required in order to capture the full error budget. This task is considerably simplified in the Bayesian approach, as each instrumental parameter is defined directly from the data themselves.

[4] https://healpix.jpl.nasa.gov

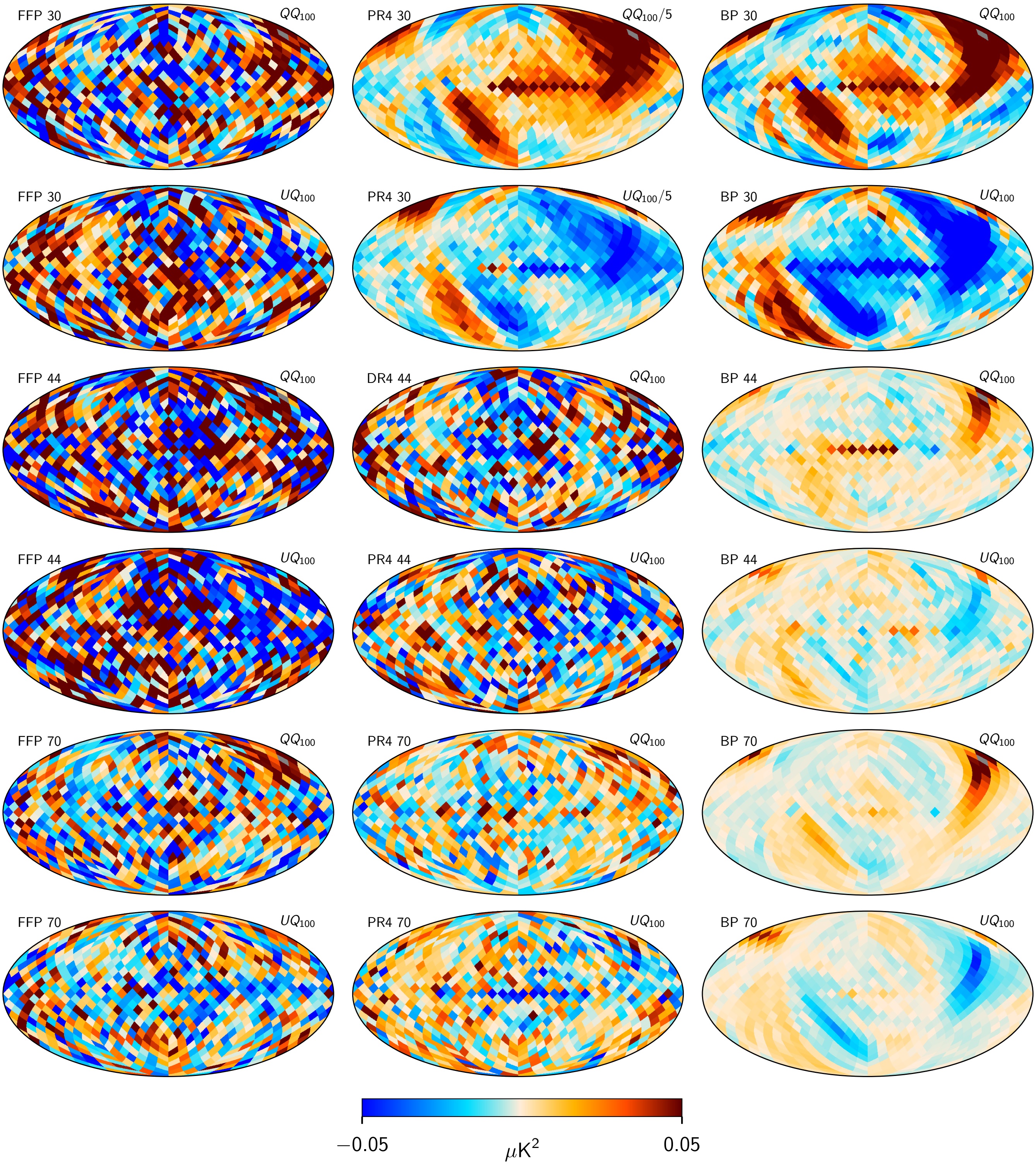


Fig. 2. Single column of the low-resolution 30 (top section), 44 (middle section), and 70 GHz (bottom section) frequency channel covariance matrix, as estimated from 300 LFI DPC FFP10 frequentist simulations (left column); from 300 PR4 prior simulations (middle column); and from 3200 BeyondPlanck posterior simulations (right column). The selected column corresponds to the Stokes Q pixel number 100 marked in gray, which is located in the top right quadrant. All covariance matrices are constructed at $N_{\rm side} = 8$. Note: the *Planck* PR4 30 GHz covariance slice has been divided by a factor of 5 and thus it is even stronger than the color scale naively implies.

4. Simulation specification

Returning to the data model summary in Sect. 2, we note that the `Commander3` code described by Galloway et al. (2023a), and used by the BeyondPlanck project to perform Bayesian end-to-end analysis of the *Planck* LFI data, is able to produce both prior and posterior simulations essentially without modifications; the only question is whether the parameters used to generate the TOD, ω, are drawn from the posterior distribution, or whether they are selected from a data-independent hyper-distribution. Choosing which type of simulations to generate is thus only a matter of selecting proper initialization values in the `Commander3` parameter file.

In this paper, we demonstrate the frequentist mode of operation by generating a set of classical frequentist simulations with `Commander3`, and we then use these to validate the novel low-level processing algorithms introduced by Keihänen et al. (2023), Ihle et al. (2023), Gjerløw et al. (2023) for mapmaking, correlated noise estimation, and gain estimation, respectively.

We note that the original BeyondPlanck analysis required 670 000 CPU-hours to generate 4000 full Gibbs samples for the full LFI dataset, which took about three months of runtime to complete. In the current paper, we are primarily interested in validating the low-level algorithms themselves, and we therefore chose to consider only one year of 30 GHz observations in the following (corresponding to about 10 000 *Planck* pointing periods (PIDs), each lasting for about one hour; Planck Collaboration I 2014), rather than the full LFI dataset, and this reduces the computational cost from 169 to 2.5 CPU-hours per Gibbs sample (Galloway et al. 2023a). As a result, we are able to produce individual chains with 10 000 samples within a matter of days, rather than months or years, which is useful for convergence analyses. This also reduces the total volume of the TOD themselves (not including pointing, flags, etc.) from 638 GB to 22 GB, and the simulations may therefore be run on a much broader range of hardware. In fact, subsets of the following simulations have been produced on more than ten different computing systems all over the world, using both AMD and Intel processors (e.g., Intel E5-2697v2 2.7 GHz, Intel Xeon E5-2698 2.3 GHz, Intel Xeon W-2255 3.7 GHz, AMD Ryzen 9 3950X 2.2 GHz), with between 128 GB and 1.5 TB RAM per node, and using both Intel and GNU compilers[5].

Given that we will only consider low-level processing of the 30 GHz channel, we simplify the data model in Eq. (1) to

$$d^{\rm sim}_{j,t} = g_{j,t}\mathsf{P}_{tp,j}\mathsf{B}^{\rm symm}_{pp',j}\,a^{\rm cmb}_{p'} + \mathsf{B}^{\rm asymm}_{pp',j}\,s^{\rm orb}_{j,t} + +n^{\rm corr}_{j,t} + n^{\rm w}_{j,t}, \tag{15}$$

$$= s^{\rm tot}_{j,t} + n^{\rm corr}_{j,t} + n^{\rm w}_{j,t}. \tag{16}$$

Here, we only included one single sky component, namely the CMB, and we ignored sub-dominant effects such as far sidelobe corrections, 1 Hz electronic spikes, etc. As such, this configuration provides a test of the gain, noise estimation, and mapmaking parts of the full algorithm, but neglecting the component separation or cosmological parameter estimation.

The CMB sky realizations used in the following analysis are drawn from the best-fit *Planck* 2018 ΛCDM model (Planck Collaboration V 2020), using the HEALPix[6] (Górski et al. 2005) `synfast` utility. All instrumental parameters are drawn from different realizations of the BeyondPlanck ensemble presented in BeyondPlanck Collaboration (2023) and these are taken as true input values in the following.

For the noise terms, we drew a random Gaussian realization of $n_{j,t} = n^{\rm corr}_{j,t} + n^{\rm w}_{j,t}$ with the noise PSD model given in Eq. (2). This was done independently for each *Planck* pointing ID (PID) and the noise PSD parameters thus vary over time with the same structure as the real observations.

[5] The research presented in this paper was undertaken as a part of the Master- and PhD-level course called "AST9240 – Cosmological component separation" in 2021 at the University of Oslo, and individual students produced and analyzed simulations in their home institutions.

[6] `http://healpix.jpl.nasa.gov`

5. Validation of low-level processing algorithms

To validate the noise and gain estimation and mapmaking steps in `Commander3`, we analyzed the prior simulations described above with the same Bayesian framework as used for the main BeyondPlanck processing and we compared the output marginal posterior distributions with the known true inputs. To quantify both biases and the accuracy of the uncertainty estimates, we adopted the following normalized residual,

$$\delta_\omega = \frac{\mu_\omega - \omega^{\rm in}}{\sigma_\omega}, \tag{17}$$

where μ_ω and σ_ω are the posterior mean and standard deviation for parameter ω. For a truly Gaussian posterior distribution with no bias and perfect uncertainty estimation, this quantity should be distributed according to a standard normal distribution with zero mean and unit variance, $N(0,1)$, while a non-zero value of δ indicates a bias measured in units of σ. It is of course important to note that the full data model in Eq. (1) is highly non-linear due to the presence of the gain. Therefore, the deviations from $N(0,1)$ at some level are fully expected, in particular for signal-dominated quantities. Still, we find that δ serves as a useful quality monitor.

Unless otherwise noted, the main results presented in the following are derived from a single Markov chain comprising 10 000 samples. Where it proves useful for convergence and mixing assessment, we also used shorter and independent chains, typically with 1000 samples in each chain.

5.1. Markov auto-correlations

We are also interested in studying the statistical properties of individual Markov chains in terms of correlation lengths, degeneracies, and convergence. We define the Markov chain auto-correlation for a given chain as:

$$\rho_\omega(\Delta) = \left\langle \left(\frac{\omega^i - \mu_\omega}{\sigma_\omega}\right)\left(\frac{\omega^{i+\Delta} - \mu_\omega}{\sigma_\omega}\right)\right\rangle, \tag{18}$$

where i denotes Gibbs sample number, and Δ is a chain lag parameter that denotes the sample separation.

Figure 3 shows the auto-correlation for a typical set of parameters. The top four panels display: (1) a single CMB map pixel (in T, Q, and U); (2) a single correlated noise map pixel (in T, Q, and U); (3) the CMB temperature quadrupole moment, $a_{2,0}$; and (4) the gain for a single PID. These all have relatively short correlation lengths, which indicates that we are likely to produce robust results for these parameters.

In contrast, the parameters in the bottom four panels have very long correlation lengths, and these correspond to the four correlated noise PSD parameters within a single PID; σ_0, $f_{\rm knee}$, α, and $A_{\rm p}/\sigma_0$. As discussed by Ihle et al. (2023), the introduction of the log-normal noise term greatly increases degeneracies and correlations among these parameters as compared to a standard $1/f$ noise profile, and this makes a proper estimation of these parameters much more expensive. However, it is also important to note that this is only a challenge regarding the estimation of the individual noise PSD parameters. In fact, the full PSD as a function of frequency, $P_n(f)$, is insensitive to these degeneracies and that function is the only thing that is actually propagated to the rest of the system. This explains why the long correlations seen in the lower half of the plot do not excite long correlations also among the (far more important) parameters in the top half of the plot.

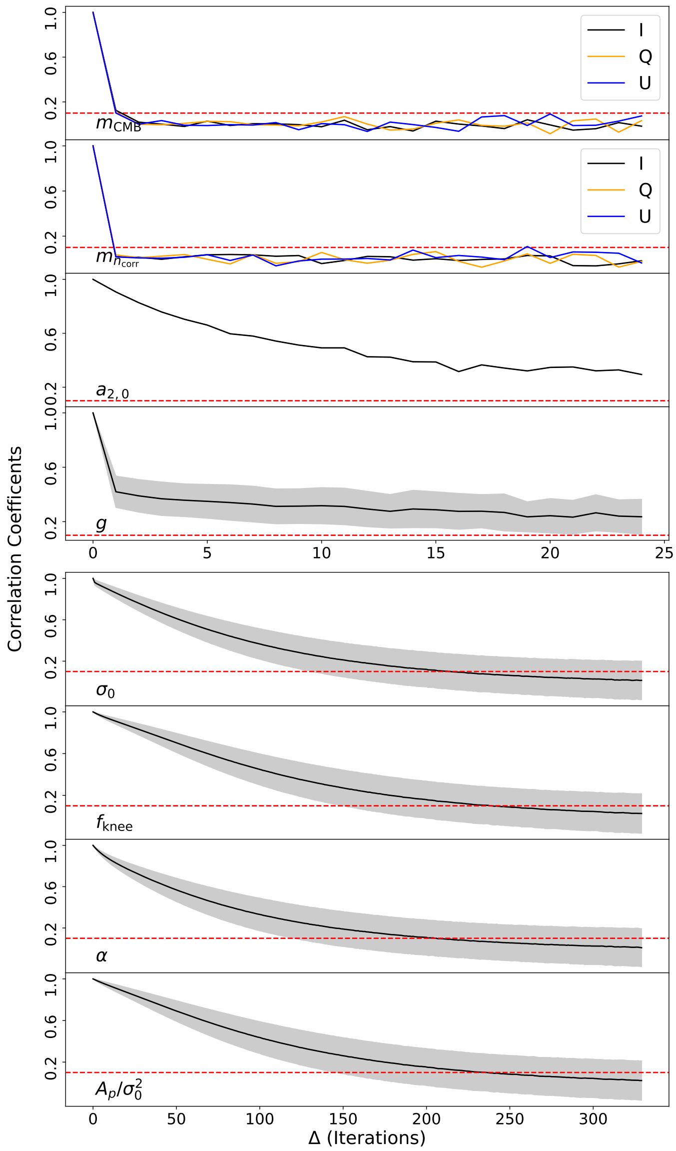


Fig. 3. Auto-correlation function, ρ, for selected parameters in the model, as estimated from a single chain with 10 000 samples. From top to bottom, the various panels show (1) one pixel value of the CMB component map $m_{\rm CMB}$; (2) one pixel of the correlated noise map $m_{n_{\rm corr}}$; (3) the temperature quadrupole moment, $a_{2,0}$; (4) the PID-averaged total gain g; and (5)–(8) the PID-averaged noise PSD parameters σ_0, $f_{\rm knee}$, α, and $A_{\rm p}/\sigma_0$. In panels with multiple lines, the various colors show Stokes T, Q, and U parameters. In panels with gray bands, the black line shows results averaged over all PIDs, and the band shows the 1σ variation among PIDs. The dashed red line marks a correlation coefficient of 0.1, which is used to define the typical correlation length of each parameter.

In fact, the single most important parameter in the entire system is the CMB map, shown in the first (for individual pixels) and third (for the quadrupole moment, $a_{2,0}$) panels. Indeed, the correlation length is very short or even non-existent for single pixels. This is primarily due to the fact that this map is strongly dominated by white noise on a single-pixel scale for the setup we consider here. As seen in the third panel, the same does not hold true for the quadrupole moment, in which case the correlation is in fact higher than 0.3 at a lag of $\Delta = 25$. The main driver for this is the gain, as shown in the fourth panel. While the gain is dominated by white noise on short time-scales (as seen by the quick drop-off between lags of 1 and 2), there is a slow drift at higher lags. This is caused by a partial degeneracy between the CMB map (which acts as a calibration source in this framework, anchored by the orbital dipole) and the overall gain. In the real BEYONDPLANCK analysis, this degeneracy is mitigated to a large extent by analyzing all LFI channels jointly, and also by including WMAP observations to break important low-ℓ polarization degeneracies (Gjerløw et al. 2023; Basyrov et al. 2023). Still, even with those additions, there are important long-term drifts in the largest CMB temperature scales, and these have non-negligible consequences for the statistical significance of low-ℓ CMB anomalies (Colombo et al. 2023).

5.2. Posterior distribution overview

Next, to build the intuition regarding the full set of recovered parameters, we show in Fig. 4 the marginal 1D and 2D posterior distributions for a small set of parameters for two different PIDs. In each panel, the true input values are shown as dashed lines. The bottom triangle (blue) show posterior results for one well-behaved PID with good goodness-of-fit statistics, while the top triangle (orange) shows a less well-behaved case in which the true input values are at the edge of recovered distributions. Together, these two cases represent the majority of all PIDs in terms of overall behaviour.

Overall, the true input parameters are recovered reasonably well in most cases. One of the parameters that is not as well recovered is the white noise amplitude, σ_0. This parameter is a special case due to the sampling algorithm currently used in the BEYONDPLANCK pipeline. As described by Ihle et al. (2023), σ_0 is currently determined as the standard deviation of all pairwise differences between neighboring time samples divided by $\sqrt{2}$. While this is a commonly used technique in radio astronomy to derive an estimate of the white noise that is highly robust against unmodeled systematic errors, it does not correspond to a proper sample from the true conditional distribution $P(\sigma_0 \mid \boldsymbol{d}, \boldsymbol{g}, \ldots)$. In particular, this approach underestimates the true fluctuations of σ_0, which in turn results in the overall uncertainties being slightly underestimated. This is one of several examples in the pipeline in which robustness to systematic effects comes at a cost of statistical rigor. At the same time, it is important to note that the absolute white noise level is in general very well determined in these data (Ihle et al. 2023) and a slight under-estimation of the uncertainty in σ_0 has little practical impact on other parameters in the model.

Looking more broadly at the 2D distributions in this figure, we see that the parameters are split naturally into two groups, defined by the short and long correlation lengths discussed above. That is to say, the CMB, correlation noise, and gain parameters generally exhibit more symmetric distributions than the noise PSD distributions, which are highly correlated and non-Gaussian. Once again, this reflects the internal degeneracies among the noise PSD parameters.

To further illustrate the impact of the slow convergence rate for several of these parameters, Fig. 5 shows four partial chains, each with only 1000 samples, for a sub-set of these parameters. Once again, we see that the input values are reasonably well recovered for most cases, but each colored subdistribution only covers a modest part of the full posterior volume.

5.3. Gain validation

In going into greater detail with respect to individual parameters, we show in Fig. 6 a subset of the estimated gain as a function

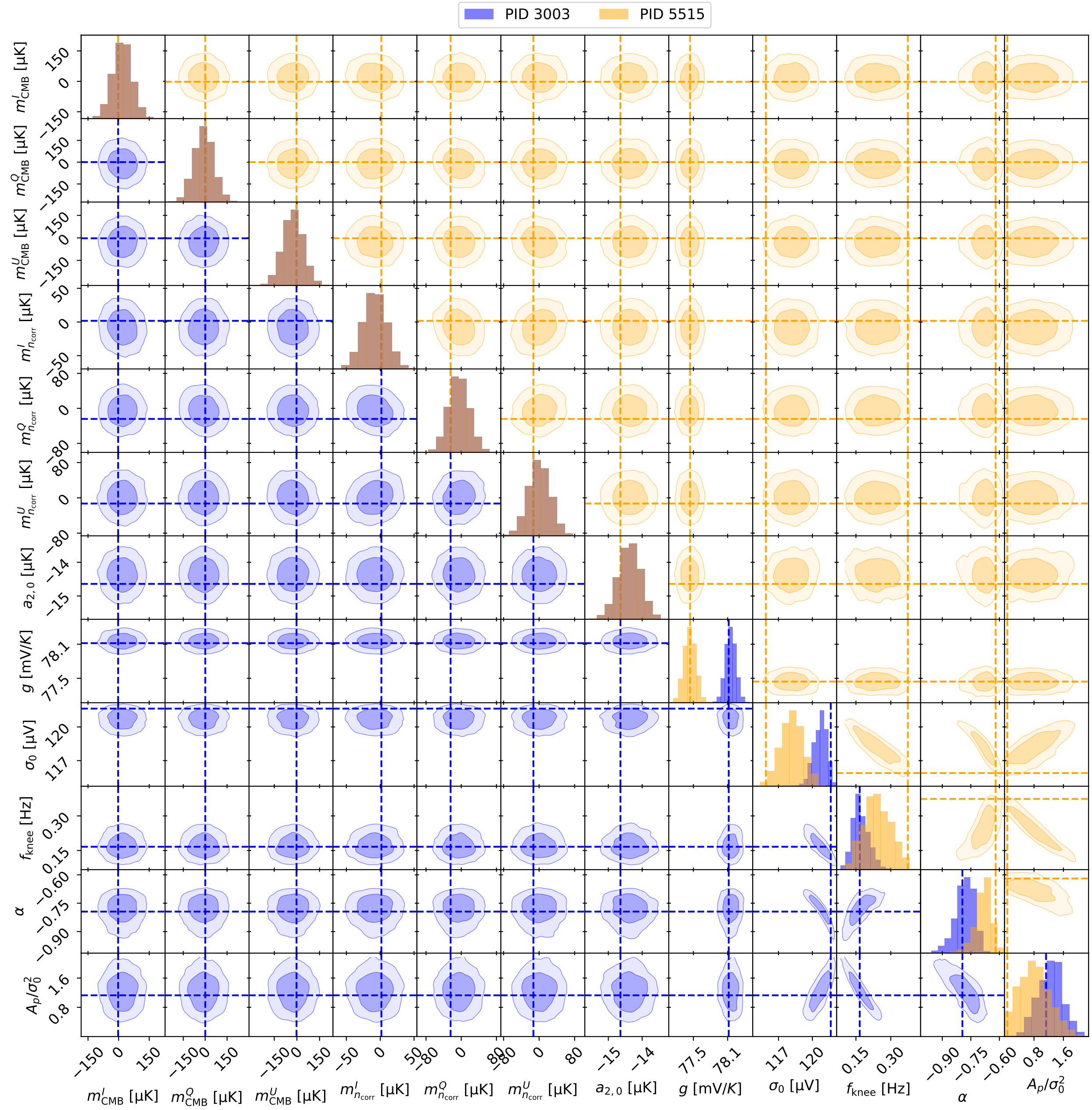


Fig. 4. Recovered posterior distributions for a selected set of parameters from two PIDs and detectors. The contours indicate 68 and 95% confidence regions, while the dashed lines (in the respective color of the contours) show the true input value of each of the PIDs. The contours below (blue) and above (orange) the diagonal correspond to PIDs 3003 and 5515, respectively. From left to right along the horizontal axis, columns show (1)–(3) one arbitrary CMB map pixel in Stokes I, Q, and U; (4)–(6) correlated noise for the same pixel and Stokes parameters; (7) the CMB intensity quadrupole amplitude $a_{2,0}$; (8) gain g; and (9)–(12) the four correlated noise parameters, $\xi^n = \{\sigma_0, f_{\rm knee}, \alpha, A_{\rm p}\}$. Note: the 1D histograms of the first seven parameters are completely overlapping since these parameters are independent of PID.

of Gibbs iteration for four selected PIDs, that is: one for each radiometer. The red lines show the true input values. Here, we visually observe the same behavior as discussed above; on short time scales, these trace plots are dominated by random fluctuations, while on long time-scales, there are still obvious significant drifts.

Figure 7 compares the estimated gain (blue bands) with the known input (red curves) as a function of PID. The width of the blue bands indicates the $\pm 1\sigma$ confidence region. At least at a visual level, the two curves agree well, without any obvious evidence of systematic biases, and the uncertainties appear reasonable. These observations are made more quantitative in Fig. 8, which shows histograms of normalized residuals, δ_g, over all PIDs. Red lines indicate the standard Gaussian $N(0, 1)$ reference distribution. Once again, we see that the reconstruction appears good, as the nominal bias is (at most) 0.36σ, and the maximum posterior width is 1.36σ. From the shape of the histograms, it is also clear that a significant fraction of these variations is due to the Monte Carlo sample variance from the long gain-correlation lengths. Once again, we note that such deviations will decrease as the number of frequency bands included in the analysis increases, since the Solar CMB dipole, which is the main calibrator, will be much better constrained with more observations. The actual gain-correlation lengths found for

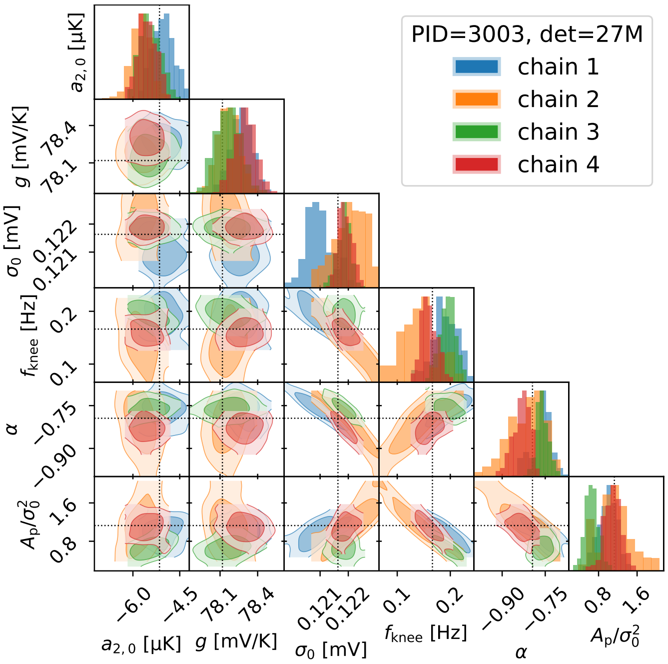


Fig. 5. Comparison of partial posteriors distributions from multiple short chains for the quadrupole amplitude, $a_{2,0}$, the gain, g, white noise level, σ_0, knee frequency, $f_{\rm knee}$, correlated noise spectral index α, and log-normal noise amplitude, $A_{\rm p}$. Each chain consists of 1000 samples. The posterior contours only span the range of the underlying samples, thereby making some of them appear open.

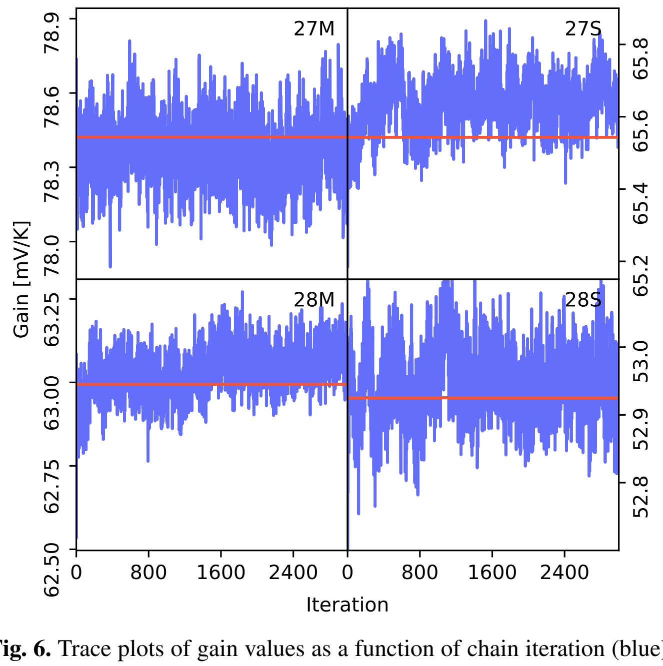


Fig. 6. Trace plots of gain values as a function of chain iteration (blue) compared to their input values (red) for selected PIDs, in order from left to right: 349, 9847, 4298, and 1993.

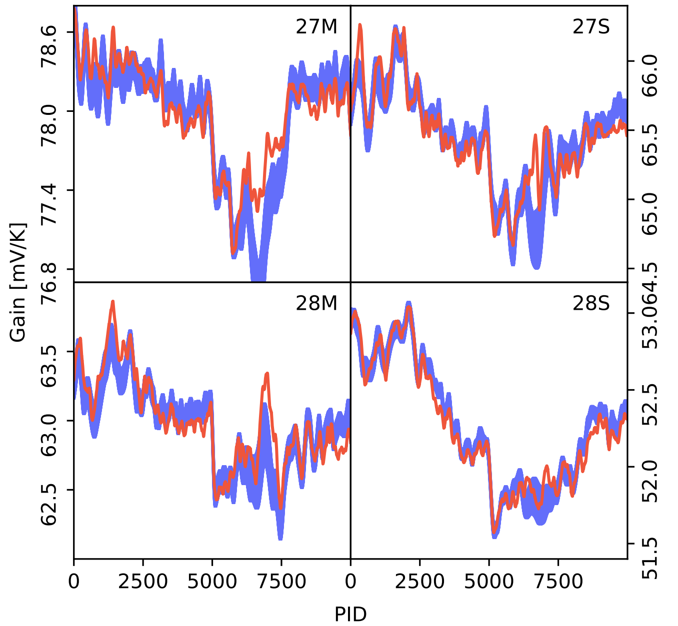


Fig. 7. Input gain values (red) over-plotted on the output gain values (blue). The width of the blue line indicates the sample standard deviation of the PID in question.

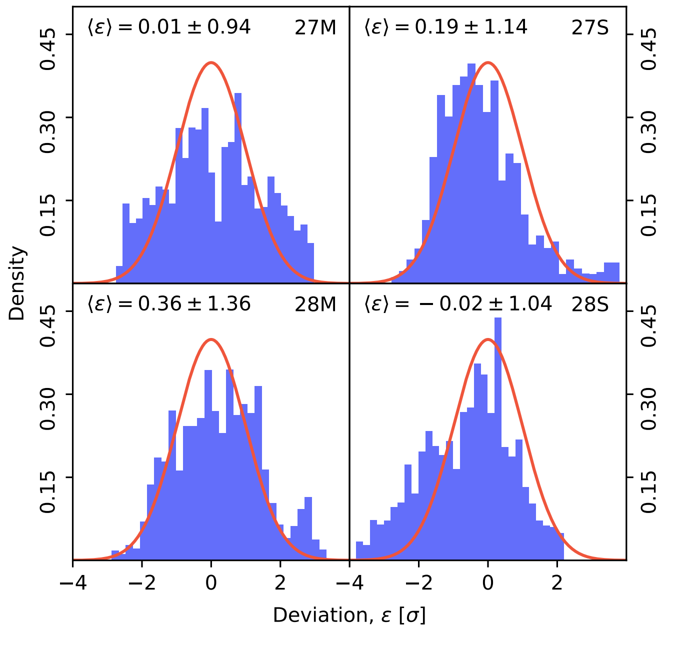


Fig. 8. Aggregate standard deviation normalized differences between the gain sample mean and the input gain values. For each PID t and detector i we calculate $(g^{\rm in}_{t,i} - \overline{g}_{t,i})/\sigma_{t,i}$, where $g^{\rm in}$ is the input gain, $\overline{g}$ is the mean sample value and σ is the sample standard deviation. We then aggregate all of these values into the appropriate histogram. The red lines are ideal Gaussian distribution for comparison. Each subplot also lists the aggregate deviation from the expected mean of 0 within the error bounds.

the real BEYONDPLANCK analysis are shown by Gjerløw et al. (2023) and they are notably shorter than those of this reduced simulation.

5.4. Correlated noise posterior validation

Next, we turn to the correlated noise component, and we start with the specific noise realization, $\boldsymbol{n}_{\rm corr}$ (the correlated noise PSD parameters is discussed separately in Sect. 5.6). To simplify the visualization, we binned the correlated noise TOD into a sky map, as illustrated in Fig. 9. The top-left panel shows the true input correlated noise map (temperature component only), while the top-right panel shows the corresponding posterior mean (output) map. The bottom-left panel shows the posterior standard

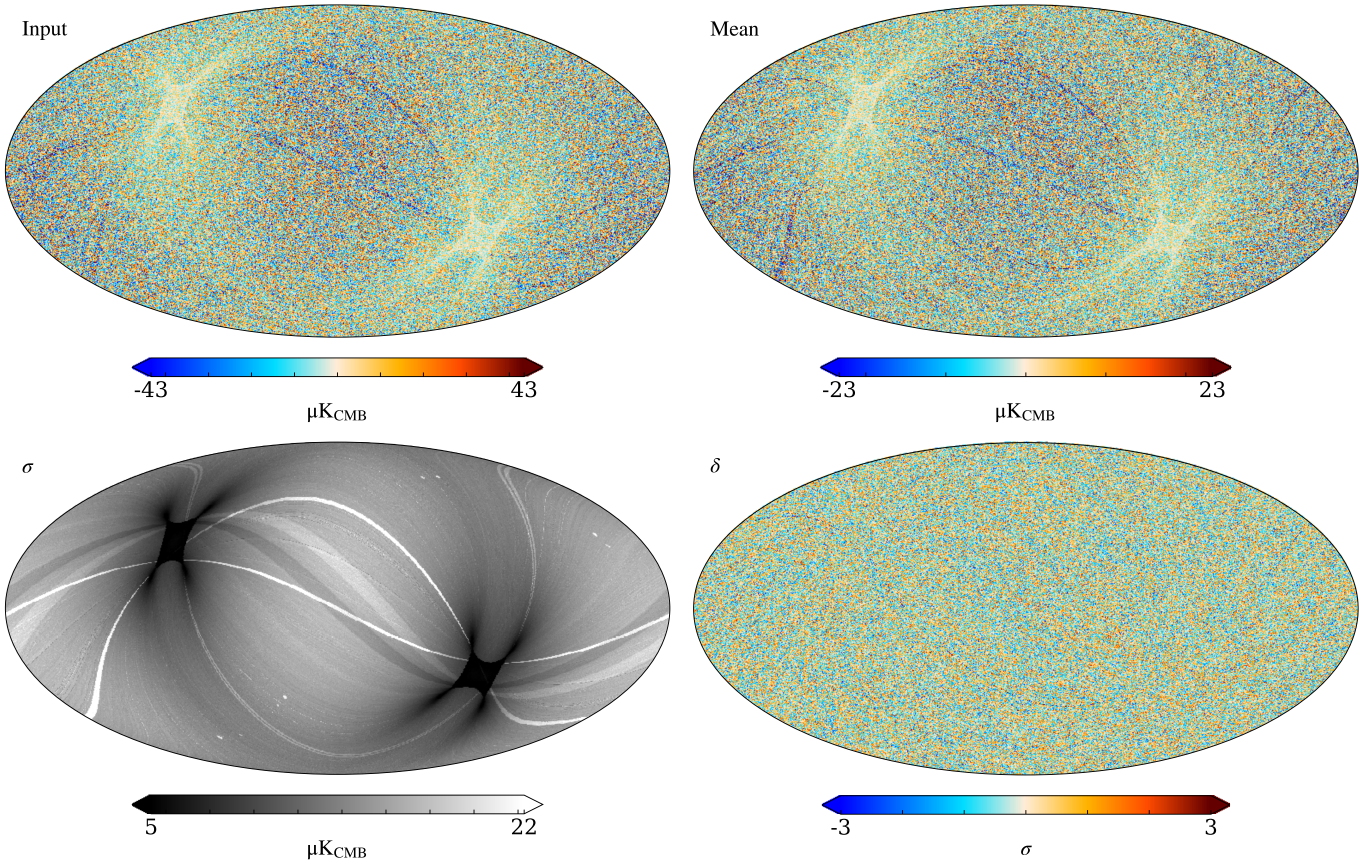


Fig. 9. Pixel space comparison of reconstructed correlated noise maps in temperature. Top left: true input realization. Top right: estimated posterior mean (output) map. Bottom left: estimated posterior standard deviation map. Bottom right: normalized residual in units of standard deviations.

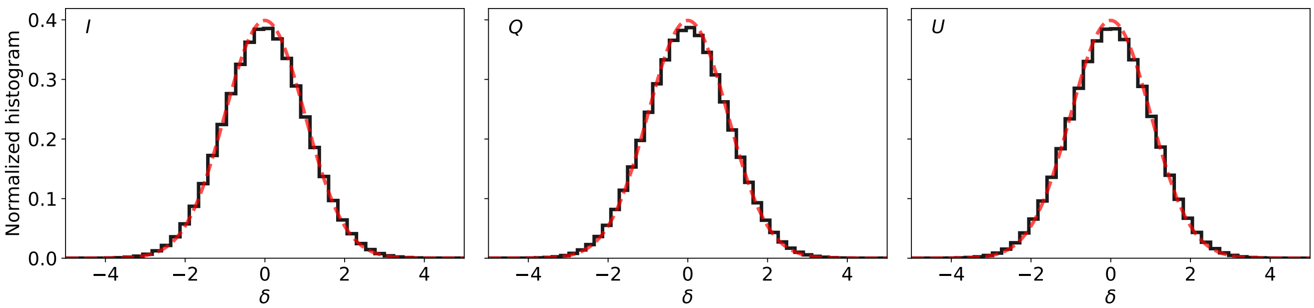


Fig. 10. Histograms of normalized correlated noise residuals, δ for each Stokes parameters (black distributions). For comparison, the dashed red line shows a standard $N(0, 1)$ distribution.

deviation per pixel and the bottom-right panel shows the normalized residual, $\delta_{\rm corr}$.

A visual inspection of the simulation input and posterior mean correlated noise maps indicates no obvious differences. In fact, the normalized residual map in the bottom right panel of Fig. 9 appears fully consistent with white noise. Once again, this observation is quantified more accurately in Fig. 10, where we compare the histogram of $\delta_{\rm corr}$ over all pixels with the usual $N(0, 1)$ distribution for each of the three Stokes parameters. In each case, the agreement is excellent.

5.5. CMB map validation

Figures 11 and 12 show similar plots for the CMB sky map component. Once again, the normalized residual in the bottom right panel appears fully consistent with white noise over most of the sky. However, this time, we actually see a power excess in $\delta_{\rm CMB}$ around the Ecliptic poles. These features correspond to regions of the sky that are particularly deeply observed by the *Planck* scanning strategy (Planck Collaboration I 2014). As a result of these deep measurements, the white noise in these regions is very low, and the total error budget per pixel is far more sensitive to the non-linear contributions in the system, in particular the coupling between the gain and the Solar dipole.

This effect does of course not only apply to the Ecliptic "deep fields", but to all signal-dominated map pixels at some level, and it therefore also applies to the full-sky CMB map in temperature. This statement is made more quantitative in the left panel of Fig. 12, where we see that the temperature histogram is very slightly wider than the reference $N(0, 1)$ distribution. To be

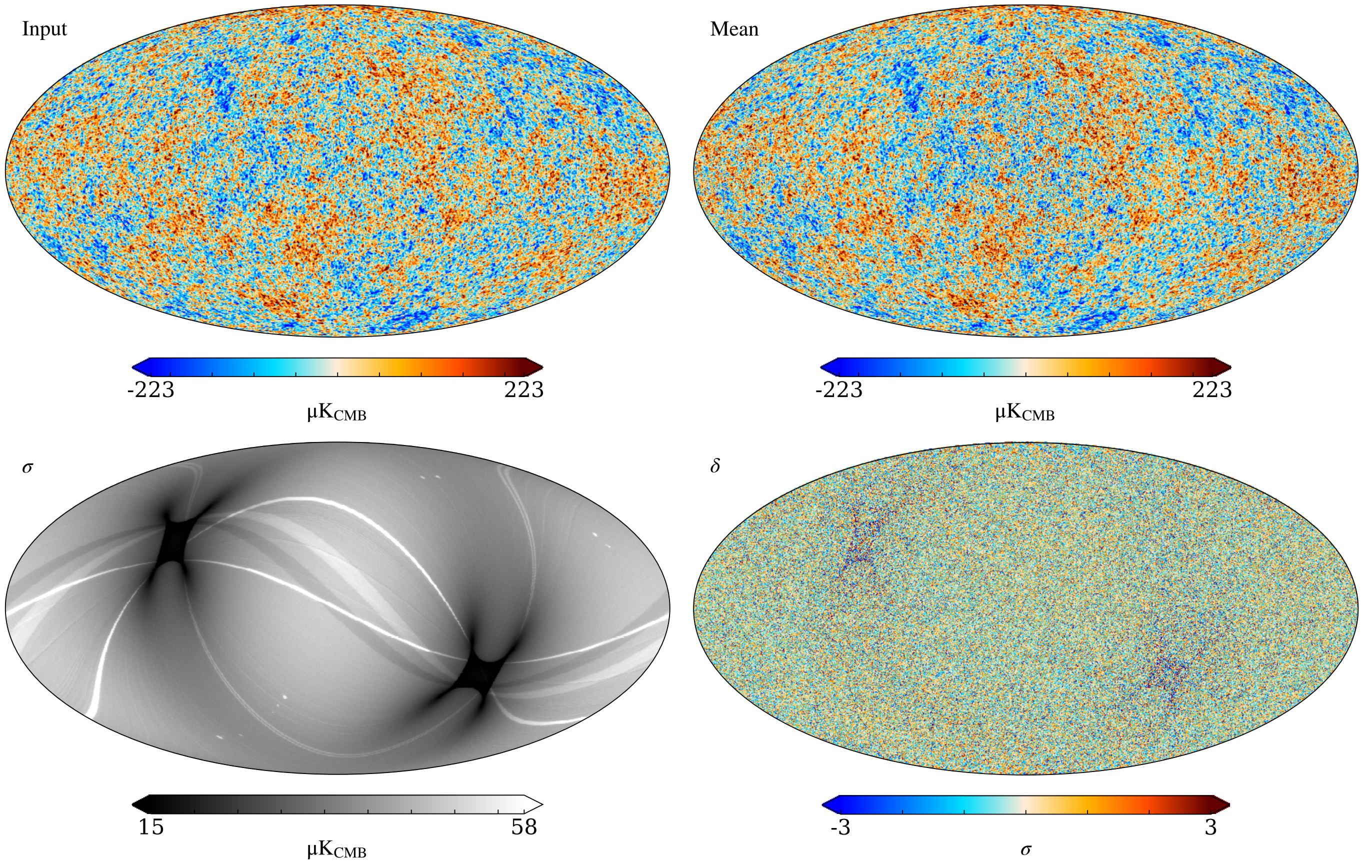


Fig. 11. Same as Fig. 9, but for the CMB intensity component.

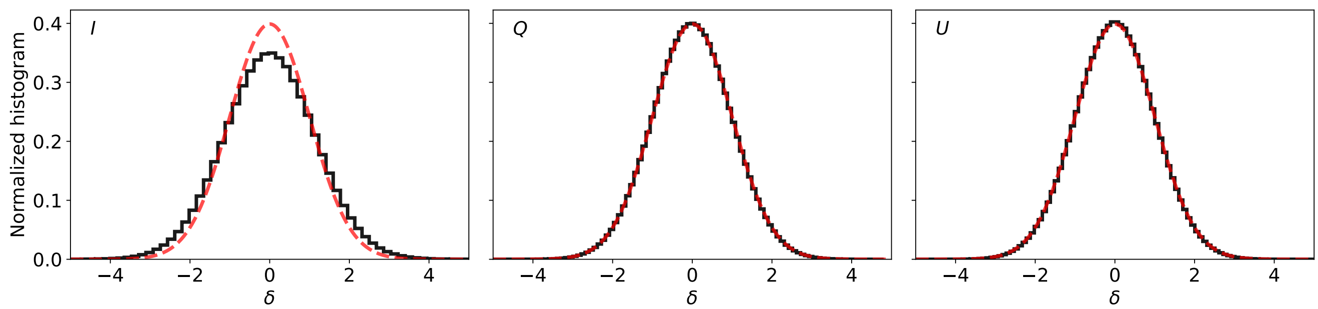


Fig. 12. Histograms of normalized CMB intensity residuals, δ_{CMB} for each Stokes parameters (black distributions). For comparison, the dashed red line shows a standard $N(0, 1)$ distribution.

specific, the standard deviation of this distribution is about 1.15. At the same time, the mean of the distribution is consistent with zero, the non-linear couplings therefore do not introduce a bias, but only a higher variance. For the noise-dominated Stokes Q and U parameters, for which gain couplings are negligible on a per-pixel level, both distributions are perfectly consistent with $N(0, 1)$.

Figure 13 shows Pearson's correlation coefficients between the CMB and correlated noise components for three selected pixels. Two of the pixels, marked '1' and '2', are located along the same *Planck* scanning ring near the Ecliptic plane, where the *Planck* scanning strategy is particularly poor. The third pixel is located far away from these, and on a different scanning ring. Here, we see that correlations are very strong for Stokes parameters of the same type along the same ring, with correlation coefficients ranging between 0.5 and 0.8. These correlations are induced both by gain and correlated-noise fluctuations, which are tightly associated with the *Planck* scanning rings. Stokes parameters of different types (e.g. I and Q) are significantly less correlated, typically with anti-correlation coefficients of $\rho \lesssim -0.25$. Correlations between widely separated pixels are practically negligible in the current simulation setup, although for the real analysis, this is no longer true due to additional couplings from, for instance, astrophysical foregrounds, bandpass corrections, and sidelobes (Galloway et al. 2023b; Svalheim et al. 2023a,b; Basyrov et al. 2023; Colombo et al. 2023; Andersen et al. 2023).

5.6. Correlated noise PSD validation

Finally, we consider the noise PSD parameters, σ_0, f_{knee}, α, and A_{p}/σ_0. As already noted, these are significantly harder to

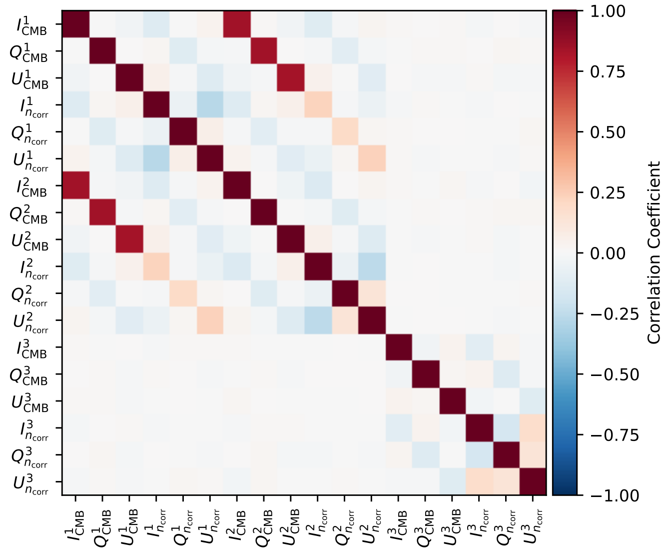


Fig. 13. Correlation matrix for selected pixel values of the CMB map, m_{CMB}, and the correlated noise map, $m_{n_{\mathrm{corr}}}$, for all three Stokes parameters I, Q, and U. Pixels 1 and 2 are selected to be neighboring pixels along the same *Planck* scanning ring and located near the Ecliptic plane, while pixel 3 is an arbitrarily selected pixel not spatially associated with the other two.

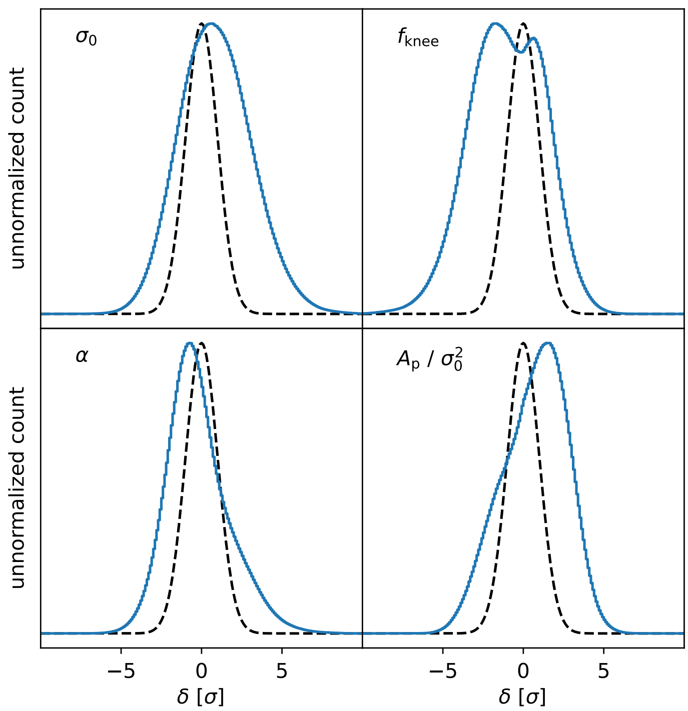


Fig. 14. Histograms of the noise parameters over all PIDs and 10 000 samples for radiometer 27M. We show the white noise level, σ_0, knee frequency, f_{knee}, correlated noise spectral index α, and log-normal noise amplitude, A_{p}. For reference, we show the standard normal distribution as a black dashed line.

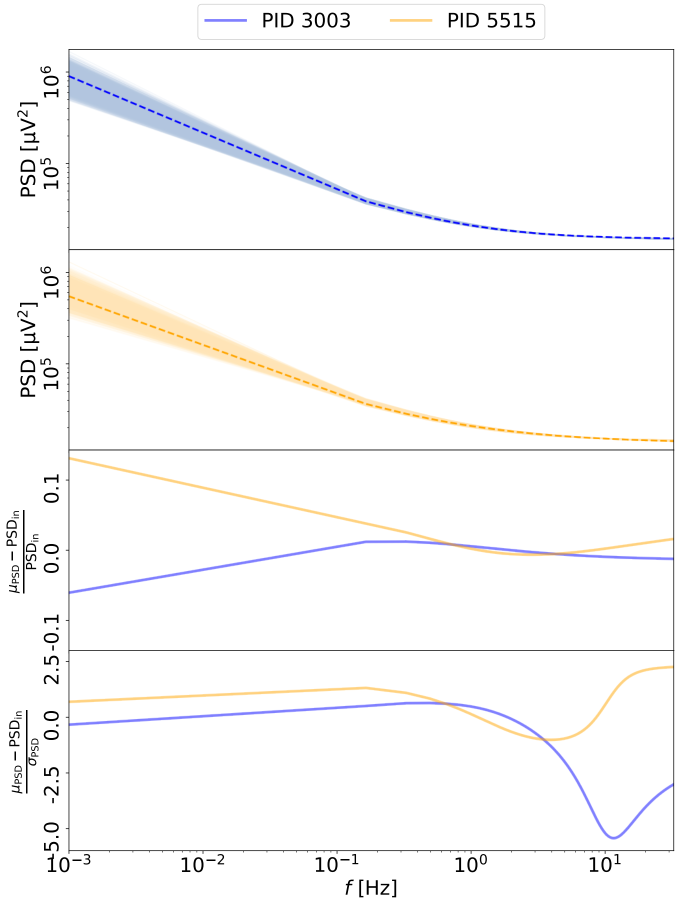


Fig. 15. Comparison of recovered correlated noise PSD in terms of the functional form, $P_n(f)$. The top two panels show results for the same PIDs as in Fig. 4. Faint lines indicate individual Gibbs samples, while the dashed lines show the true input functions. The bottom two panels show the difference between the posterior mean function and the true input as a fraction of the latter and in units of the posterior rms, respectively.

estimate individually than the previous parameters due to the strong correlation between the $1/f$ and log-normal terms in Eq. (2).

As usual, we plot the reduced residual, δ, for each parameter type in Fig. 14. In this case, we see that the posterior distributions are significantly wider than a standard Gaussian distribution, by as much as a factor of two. The distributions are also clearly non-Gaussian, with notable skewness and kurtosis. Both the excess variance and non-Gaussianity stem from the same degeneracies as discussed above and are partially due to intrinsic non-Gaussianities in the model, and partially due to incomplete Monte Carlo convergence and very long correlation lengths. On the other hand, the mean bias in these distribution is small and the estimated posterior distributions do provide a useful summary of each parameter individually.

As mentioned above, however, other parameters in the model are not sensitive to individual ξ^n values, but only to the total noise PSD, $P_{\mathrm{corr}}(f)$. This function is plotted in the top two panels of Fig. 15 for the same two PIDs and radiometers as shown in Fig. 4; the blue curves correspond to the aptly measured PID, while the orange curve corresponds to the PID with the marginal fit. Faint lines in the top two panels show individual Gibbs samples, corresponding to different combinations of ξ^n. By eye, the sampled values appear to span the true input reasonably well, although the orange line is on the lower edge of the estimated posterior distribution.

These visual observations are made more quantitative in the bottom two panels, where the third panel shows the fractional difference between the output and input PSD functions, and the fourth panel shows the same in units of standard deviation of the

PSD across Gibbs samples, σ. For the well-behaved (blue) pixel, we see that the posterior mean matches the true input everywhere to within a few percent; in units of standard deviations, this is typically less than 2.5σ for most of the region, except at frequencies above 10 Hz, where the estimated standard deviation is very small due, and the underestimation of the uncertainty in σ_0 becomes noticeable. For the less well-behaved case, the recovered PSD is within 2σ at all frequencies in units of standard deviations, or within 5 % otherwise. Overall, the PSD itself is recovered very well in both cases in absolute terms.

6. Conclusions

End-to-end time-ordered simulations play a key role in estimating both biases and uncertainties for current and future CMB experiments. To date, no other practical method has been able to account for the full and rich set of systematic errors that affect modern high-precision measurements.

As detailed in BeyondPlanck Collaboration (2023) and its companion papers, the BeyondPlanck project has implemented a new approach to end-to-end CMB analysis in which a global parametric model is fitted directly to the time-ordered data, allowing for joint estimation of instrumental, astrophysical, and cosmological parameters with true end-to-end error propagation. This approach relies strongly on a sampling algorithm called Gibbs sampling, which allows the user to draw joint samples from a complex posterior distribution. Each of these Gibbs samples correspond essentially to one end-to-end TOD simulation, similar to those produced by classical CMB simulation pipelines, such as the *Planck* full focalplane (FFP; Planck Collaboration XII 2016) simulations.

The fundamental difference between these two simulation pipelines lies in how to define the input parameters used to generate the simulation. In the BeyondPlanck approach, all parameters are constrained directly from the true data and correspond as such to samples drawn from the full joint posterior distribution. In contrast, traditional pipelines use parameters that are a mixture of data-constrained and data-independent parameters. Typical examples of the former include the CMB Solar dipole and Galactic foregrounds, while typical examples of the latter include CMB anisotropies and instrumental noise. In this paper, we call the two types of simulations for posterior- or prior-based, respectively, indicating whether (or not) they are conditioned on true data.

The difference between these two types of simulations has direct real-world consequences for what applications each simulation type is suitable for. As was first argued by Basyrov et al. (2023), this may be intuitively understood through the following line of reasoning. Supposing we are looking to construct a new end-to-end simulation for a given experiment. Among the first decisions that needs to be made concerns the CMB Solar dipole, answering the question of whether this should correspond to the true dipole or whether it should have a random amplitude and direction. If it is chosen randomly, then the hot and cold spots in the correlation matrices shown in Fig. 2 in this paper will appear at random positions on the sky, and eventually be washed out in an ensemble average. In practice, all current pipelines adopt the true CMB Solar dipole as an input. The next question is related to the type of Galactic model that should be used. Once again, if this is selected randomly, then the Galactic plane will move around on the sky from realization to realization. In practice, all current pipelines adopt a model of the true Galactic signal as an input.

The third question considers which CMB anisotropies should be used. At this point, all pipelines prior to BeyondPlanck have in fact adopted random CMB skies drawn from a theoretical ΛCDM model. This has two main effects: on the one hand, in the same way that randomizing the CMB dipole signal would average out any coherent correlations between the sky signal and the gain, randomizing the CMB anisotropies also average out, and non-linear correlations between these structures and the instrumental parameters are not accounted for. On the other hand, the resulting simulations do actually include so-called cosmic variance, that is, for the scatter between individual CMB realizations.

Finally, the same question apply to all the instrumental parameters, perhaps most notably correlated noise and gain fluctuations: we ask whether these ought to be constrained by the real data, or whether they should be drawn randomly from a laboratory-determined hyper-distribution.

It is important to stress that none of these four questions have a "right" or "wrong" answer. However, whatever choice is made, it will have direct consequences for what correlation structures appear among the resulting simulations and, thus, also for the sorts of applications they are suitable for. In particular, if the primary application is traditional frequentist model testing (e.g., asking whether the CMB sky is Gaussian and isotropic), then it is critical to account for cosmic variance among the CMB realizations. For those applications, we must choose data-independent CMB inputs in order to capture the full uncertainties and the appropriate choice are frequentist data-independent simulation inputs.

If, on the other hand, the main application is traditional parameter estimation, for instance, to constrain the ΛCDM model, then it is key to properly estimate the total CMB uncertainty per-pixel on the sky. In this case, it is critical to properly model all non-linear couplings between the actual sky signal, the true gain, the true correlated noise, and so on. In this case, the appropriate choices are posterior-based data-dependent simulation inputs.

In this paper, we note that the novel `Commander3` software is able to produce both prior and posterior simulations simply by adjusting the inputs that are used to initialize the code. While the posterior simulation process has been described in detail in most of the BeyondPlanck companion papers, in the current paper, we present a first application of the frequentist mode of operation by producing a data-independent time-ordered simulation corresponding to one year of 30 GHz data. We we then used this to validate three important low-level steps in the full BeyondPlanck Gibbs samples, namely, gain estimation, correlated noise estimation, and mapmaking. In doing so, we find that the recovered posterior distribution matches the true input parameters well.

Acknowledgements. We thank Prof. Pedro Ferreira and Dr. Charles Lawrence for very useful suggestions, comments and discussions. We also thank the entire *Planck* and WMAP teams for invaluable support and discussions, and for their dedicated efforts through several decades without which this work would not be possible. This research has partially been carried out within the master- and PhD-level course "AST9240 – Cosmological component separation" at the University of Oslo, and support for this has been provided by the Research Council of Norway through grant agreement no. 274990. The BeyondPlanck Collaboration has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement numbers 776282, 772253, and 819478. In addition, the collaboration acknowledges support from ESA; ASI, CNR, and INAF (Italy); NASA and DoE (USA); Tekes, AoF, and CSC (Finland); RCN (Norway); ERC and PRACE (EU). L. T. Hergt was supported by a UBC Killam Postdoctoral Research Fellowship. The work of Tamaki Murokoshi was supported by MEXT KAKENHI Grant Number 18H05539 and the Graduate

Program on Physics for the Universe (GP-PU), Tohoku University. Part of the work by F. Rahman was carried out using the Nova and Coolstar clusters at the Indian Institute of Astrophysics, Bangalore. The work of K. S. F. Fornazier, G. A. Hoerning, A. Marins and F. B. Abdalla was developed in the Brazilian cluster SDumont and was supported by São Paulo Research Foundation (FAPESP) grant 2014/07885-0, National Council for Scientific and Technological (CNPQ), Coordination for the Improvement of Higher Education Personnel (CAPES) and University of São Paulo (USP). Some of the results in this paper have been derived using the healpy (Zonca et al. 2019) and HEALPix (Górski et al. 2005) packages. This work made use of Astropy (http://www.astropy.org) (Astropy Collaboration 2013, 2018, 2022).

References

Andersen, K. J., Herman, D., Aurlien, R., et al. 2023, A&A, 675, A13 (BeyondPlanck SI)
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Astropy Collaboration (Price-Whelan, A. M., et al.) 2022, ApJ, 935, 167
Basyrov, A., Suur-Uski, A.-S., Colombo, L. P. L., et al. 2023, A&A, 675, A10 (BeyondPlanck SI)
Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20
BeyondPlanck Collaboration (Andersen, K. J., et al.) 2023, A&A, 675, A1 (BeyondPlanck SI)
Colombo, L. P. L., Eskilt, J. R., Paradiso, S., et al. 2023, A&A, 675, A11 (BeyondPlanck SI)
Eriksen, H. K., O'Dwyer, I. J., Jewell, J. B., et al. 2004, ApJS, 155, 227
Galloway, M., Andersen, K. J., Aurlien, R., et al. 2023a, A&A, 675, A3 (BeyondPlanck SI)
Galloway, M., Reinecke, M., Andersen, K. J., et al. 2023b, A&A, 675, A8 (BeyondPlanck SI)
Gelman, A., Vehtari, A., Simpson, D., et al. 2020, ArXiv e-prints [arXiv:2011.01808]
Gerakakis, S., Brilenkov, M., Ieronymaki, M., et al. 2023, OJA, 6, 10
Gjerløw, E., Ihle, H. T., Galeotta, S., et al. 2023, A&A, 675, A7 (BeyondPlanck SI)
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Ihle, H. T., Bersanelli, M., Franceschet, C., et al. 2023, A&A, 675, A6 (BeyondPlanck SI)
Keihänen, E., Suur-Uski, A.-S., Andersen, K. J., et al. 2023, A&A, 675, A2 (BeyondPlanck SI)
LiteBIRD Collaboration (Allys, E., et al.) 2023, PTEP, 2023, 042F01
Planck Collaboration I. 2014, A&A, 571, A1
Planck Collaboration I. 2016, A&A, 594, A1
Planck Collaboration XII. 2016, A&A, 594, A12
Planck Collaboration I. 2020, A&A, 641, A1
Planck Collaboration II. 2020, A&A, 641, A2
Planck Collaboration V. 2020, A&A, 641, A5
Planck Collaboration Int. LVII., 2020, A&A, 643, A42
Reinecke, M., Dolag, K., Hell, R., Bartelmann, M., & Ensslin, T. A. 2015, Astrophysics Source Code Library [record ascl:1505.032]
Svalheim, T. L., Zonca, A., Andersen, K. J., et al. 2023a, A&A, 675, A9 (BeyondPlanck SI)
Svalheim, T. L., Andersen, K. J., Aurlien, R., et al. 2023b, A&A, 675, A14 (BeyondPlanck SI)
Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, Phys. Rev. D, 70, 083511
Watts, D. J., Galloway, M., Ihle, H. T., et al. 2023, A&A, 675, A16 (BeyondPlanck SI)
Zonca, A., Singer, L., Lenz, D., et al. 2019, J. Open Source Softw., 4, 1298

[1] Institute of Theoretical Astrophysics, University of Oslo, Blindern, Oslo, Norway
[2] Instituto de Física, Universidade de São Paulo, C.P. 66318, CEP:, 05315-970 São Paulo, Brazil
[3] Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z1, Canada
[4] Astronomical Institute, Tohoku University, Sendai, Miyagi 980-8578, Japan
[5] Indian Institute of Astrophysics, Koramangala II Block, Bangalore 560034, India
[6] Department of Physics, University of California, Berkeley, Berkeley, CA, USA
[7] Instituto Nacional de Pesquisas Espaciais, Divisão de Astrofísica, Av. dos Astronautas, 1758, 12227-010 São José dos Campos, SP, Brazil
[8] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
[9] Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa
[10] Planetek Hellas, Leoforos Kifisias 44, Marousi 151 25, Greece
[11] Dipartimento di Fisica, Università degli Studi di Milano, Via Celoria, 16, Milano, Italy
[12] INAF/IASF Milano, Via E. Bassini 15, Milano, Italy
[13] INFN, Sezione di Milano, Via Celoria 16, Milano, Italy
[14] INAF – Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, Trieste, Italy
[15] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA
[16] Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
[17] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, USA
[18] Department of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
[19] Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
[20] Computational Cosmology Center, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[21] Haverford College Astronomy Department, 370 Lancaster Avenue, Haverford, Pennsylvania, USA
[22] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany
[23] Dipartimento di Fisica, Università degli Studi di Trieste, Via A. Valerio 2, Trieste, Italy

Paper VII

Optimal bolometer transfer function deconvolution for CMB experiments through maximum likelihood mapmaking

A&A, 688, A166 (2024)
https://doi.org/10.1051/0004-6361/202450712

Astronomy & Astrophysics

Optimal bolometer transfer function deconvolution for CMB experiments through maximum likelihood mapmaking

A. Basyrov, N. O. Stutzer, J. G. S. Lunde, H. K. Eriksen, E. Gjerløw, D. J. Watts, and I. K. Wehus

Institute of Theoretical Astrophysics, University of Oslo, Blindern, Oslo, Norway
e-mail: artem.basyrov@astro.uio.no

Received 14 May 2024 / Accepted 24 June 2024

ABSTRACT

We revisit the impact of finite time responses of bolometric detectors used for deep observations of the cosmic microwave background (CMB). Until now, bolometer transfer functions have been accounted for through a two-step procedure by first deconvolving an estimate of their Fourier-space representation from the raw time-ordered data (TOD), and then averaging the deconvolved TOD into pixelized maps. However, for many experiments, including the *Planck* High Frequency Instrument (HFI), it is necessary to apply an additional low-pass filter to avoid an excessive noise boost, which leads to an asymmetric effective beam. In this paper we demonstrate that this effect can be avoided if the transfer function deconvolution and pixelization operations are performed simultaneously through integrated maximum likelihood mapmaking. The resulting algorithm is structurally identical to the `artDeco` algorithm for beam deconvolution. We illustrate the relevance of this method with simulated *Planck* HFI 143 GHz data, and find that the resulting effective beam is both more symmetric than with the two-step procedure, resulting in a sky-averaged ellipticity that is 64% lower, and an effective beam full-width-at-half-maximum (FWHM) that is 2.3% smaller. Similar improvements are expected for any other bolometer-based CMB experiments with long time constants.

Key words. instrumentation: detectors – methods: data analysis – methods: numerical – methods: statistical – cosmic background radiation – cosmology: observations

1. Introduction

During the last three decades, our understanding of the cosmic microwave background (CMB) has been revolutionized by a series of increasingly sensitive instruments (e.g., Bennett et al. 1996; Hinshaw et al. 2013; Planck Collaboration I 2014, 2016, 2020). These advances have been made possible by increased sensitivity, driven by improvements in detector technology both for coherent radiometer and incoherent bolometer detectors.

Bolometric detectors in particular have gone a long way in improving both sensitivity and the frequency range they are able to operate in (e.g., Zhao et al. 2008; Bersanelli et al. 2010; Lamarre et al. 2010; Stevens et al. 2020). One of the main characteristics of a bolometer is a finite time constant that describes its temporal response to a signal change (e.g., Planck Collaboration VII 2014). The main observational signature of a finite bolometer transfer function is an apparent smoothing of the true underlying signal along the scanning path of the instrument. Fortunately, the magnitude of this effect has diminished over time, as the bolometer detector technology has improved and the response rates have become faster. Still, this effect has to be accounted for during mapmaking in order to establish an accurate estimate of the true sky signal.

The traditional approach to account for this effect is simply to deconvolve an estimate of the bolometer transfer function from the time-ordered data (TOD), which results in an unbiased signal. A significant drawback of this method, however, is that it not only affects the sky signal, but also the noise measured by the detector, which originates at a later point in the electronic circuitry and is not affected by the bolometer time constant at all. The noise is therefore effectively amplified on short time-scales. In the original *Planck* analysis, this problem was solved by applying an extra low-pass regularization filter function that suppresses high-frequency noise (Planck Collaboration VII 2014). While using this filter does solve the noise amplification problem at high temporal frequencies, it also modifies the signal, thereby introducing an extra component to the effective beam.

In this work, we propose an alternative approach that exploits the same ideas as proposed for beam deconvolution by Keihänen & Reinecke (2012). Specifically, rather than explicitly deconvolving the beam transfer function in a pre-processing step prior to mapmaking, we integrate the deconvolution operator directly into the maximum likelihood mapmaking equation, which then is solved using a conjugate gradient method (e.g., Shewchuk 1994). This approach has several advantages. Firstly, it does not require an explicit additional noise regularization kernel, but relies on the scanning strategy itself to regularize the high-frequency noise. This method yields an unbiased estimate of the true sky signal without modifying the effective beam. Secondly, it results in significantly weaker noise correlations at high temporal frequencies. The main drawback of the method is a higher computational expense.

The rest of the paper is organized as follows. We first describe the new method in Sect. 2. We then illustrate the main points with a simple one-dimensional case in Sect. 3 in which all calculations can be performed by using dense linear algebra. Next, we consider the two-dimensional case in Sect. 4, and start by characterizing its performance for a grid of point sources. Finally, we apply the method on the simulated CMB map in Sect. 5 with properties similar to the *Planck* 143 GHz channel. We discuss the computational cost of the proposed method in Sect. 6, before concluding in Sect. 7.

2. Method

We start by assuming that the data $\boldsymbol{d}$ recorded by a bolometer may be modeled as

$$\boldsymbol{d} = \mathsf{T}\boldsymbol{s} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{s}$ is the true sky signal projected from the sky map as $\boldsymbol{s} = \mathsf{P}\boldsymbol{m}_s$; P is a pointing matrix of size $(n_{\mathrm{pix}}, n_{\mathrm{tod}})$ that maps between pixel space and time ordered data; T represents convolution with a bolometer transfer function $\mathsf{T} = \mathsf{F}^{-1}\mathsf{T}(\omega)\mathsf{F}$, where F denotes a Fourier transform; and $\boldsymbol{n}$ denotes noise. In this work, we assume that the latter only consists of zero-mean Gaussian noise, and we define its covariance matrix as $\mathsf{N} \equiv \langle \boldsymbol{n}\boldsymbol{n}^T \rangle$. However, there are many other sources of instrumental noise in the real data (e.g., Planck Collaboration VII 2016; Ihle et al. 2023) that must be taken into account in a full analysis pipeline.

Our goal now is to derive an accurate estimate of $\boldsymbol{s}$ given some observed data, and we will denote this estimate $\hat{\boldsymbol{m}}$. The traditional approach for doing this adopted by most bolometer-based experiments consists of a two-step procedure; one specific example that is particularly relevant for this paper is *Planck* HFI (Planck Collaboration VII 2014, 2016). The first step is to explicitly apply the inverse transfer function operator to the raw TOD,

$$\mathsf{T}^{-1}\boldsymbol{d} = \mathsf{T}^{-1}(\mathsf{T}\boldsymbol{s} + \boldsymbol{n}) = \boldsymbol{s} + \mathsf{T}^{-1}\boldsymbol{n}. \tag{2}$$

As long as T is non-singular, this results in an unbiased estimate of the signal $\boldsymbol{s}$. However, it also modifies the noise, $\boldsymbol{n}$. In particular, due to the shape of the transfer function (as shown in Fig. 1), this inverse operator significantly boosts the noise level at high frequencies in Fourier space. To prevent the introduction of excessive noise, a common solution is to apply an additional low-pass filter function $K(\omega)$, such that the total filtered TOD reads

$$\mathsf{K}\mathsf{T}^{-1}\boldsymbol{d} = \mathsf{K}\mathsf{T}^{-1}(\mathsf{T}\boldsymbol{s} + \boldsymbol{n}) = \mathsf{K}\boldsymbol{s} + \mathsf{K}\mathsf{T}^{-1}\boldsymbol{n}. \tag{3}$$

Here we have introduced a filter operator K similar to the transfer function operator T as $\mathsf{K} = \mathsf{F}^{-1}\mathsf{K}(\omega)\mathsf{F}$.

The second step in the traditional procedure is to apply a mapmaking algorithm to this deconvolved TOD. Under the assumption of Gaussian noise, the optimal solution for this is given by the normal equations (e.g., Tegmark 1997),

$$\mathsf{P}^T\mathsf{N}^{-1}\mathsf{P}\hat{\boldsymbol{m}}_{\mathrm{trad}} = \mathsf{P}^T\mathsf{N}^{-1}\boldsymbol{d}. \tag{4}$$

In practice, this optimal mapmaking equation is often replaced with a computationally cheaper solution that does not require inversion of a full dense noise covariance matrix. In many cases N is simply approximated with its diagonal, and in that case the equation may be solved pixel-by-pixel (so-called binning). Another common solution is to apply a destriping algorithm, which accounts for large-scale noise fluctuations. In either case, $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ is sub-optimal in two respects: First, if $\mathsf{K} \neq \mathsf{I}$, then $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ is a biased estimator of $\boldsymbol{s}$. In practice, this is typically accounted for in higher-level analysis by modifying the effective instrumental beam, which then introduces significant asymmetries that couple to the scanning strategy. Second, the actual noise covariance matrix in the post-deconvolved TOD reads $\mathsf{K}\mathsf{T}^{-1}\mathsf{N}\mathsf{T}^{-1}\mathsf{K}$, but these additional terms are not accounted for in the above solution. As such, the noise weighting of $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ is also sub-optimal.

Aiming to resolve both these deficiencies, we adopt a simpler approach in this paper, and note that an unbiased and optimal estimate of $\hat{\boldsymbol{m}}$ can be obtained directly from the data model in Eq. (1) as follows,

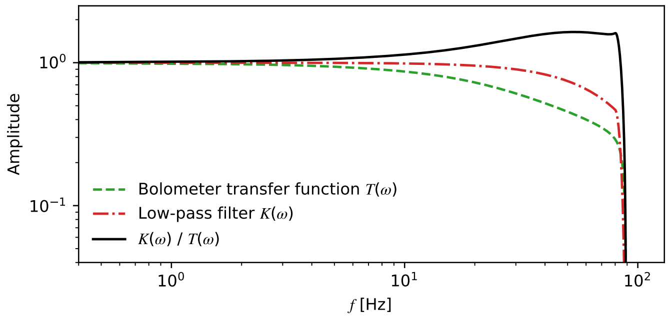


Fig. 1. Amplitudes of the *Planck* HFI bolometer transfer function $T(\omega)$ (green), the low-pass filter $K(\omega)$ employed by *Planck* HFI to keep the high-k modes from blowing up (red), and the former divided by the latter (black), which is the resulting function applied to the data in *Planck* deconvolution method. The presented functions are in absolute amplitudes. The bolometer transfer function also has a complex component.

$$\mathsf{P}^T\mathsf{T}^T\mathsf{N}^{-1}\mathsf{T}\mathsf{P}\hat{\boldsymbol{m}} = \mathsf{P}^T\mathsf{T}^T\mathsf{N}^{-1}\boldsymbol{d}. \tag{5}$$

We denote the solution of this equation $\hat{\boldsymbol{m}}_{\mathrm{MLE}}$, where MLE is short of maximum likelihood estimate. The equation involves the transpose of T, which may be written as $\mathsf{T}^T = \mathsf{F}^{-1}\mathsf{T}^*(\omega)\mathsf{F}$, where $\mathsf{T}^*(\omega) = T^*(\omega)$ is the complex conjugate of the transfer function $T(\omega)$.

We use a standard preconditioned conjugate gradient method (CG; Shewchuk 1994) to solve Eq. (5), and we find that a simple diagonal preconditioner of the form

$$\mathsf{M} = \mathsf{P}^T\mathsf{P} = \sum_{tt'} \mathsf{P}_{tp}\mathsf{P}_{t'p} \tag{6}$$

results in a speed-up of almost a factor of 10 compared to no preconditioning. This algorithm is conceptually identical to the `artDeco` algorithm introduced by Keihänen & Reinecke (2012) for asymmetric beam deconvolution, the main difference being that our T operator is computationally much cheaper than their asymmetric beam operator B.

3. One-dimensional toy model: Intuition

In general, our own primary motivation for this line of work lies in a future reanalysis of the *Planck* HFI observations. We therefore adopt the HFI 143-5 bolometer transfer function $T(\omega)$ and low-pass filter $K(\omega)$ as an explicit test case, which is shown in Fig. 1. The goal of this and the following two sections is to compare the performance of the traditional and the optimal methods in various settings for this case. In fact, most of the key algebraic points can be easily demonstrated and visualized through a simple one-dimensional case in which all matrix operations can be solved quickly by brute-force methods

In this first example, we define our true input sky map to consist of an array with 200 one-dimensional pixels. This signal map is then scanned by a simple sinusoidal scanning strategy $\theta = \sin(2\pi f t)$, where $f = 0.2$ Hz and the sampling rate is $f_{\mathrm{samp}} = 180.3737$ Hz, similar to *Planck* HFI 143 GHz sampling rate. The resulting signal-only TOD has length $n_{\mathrm{tod}} = 10\,000$ and is then convolved with the *Planck* transfer function shown in Fig. 1. Finally, white Gaussian noise is added. We then solve

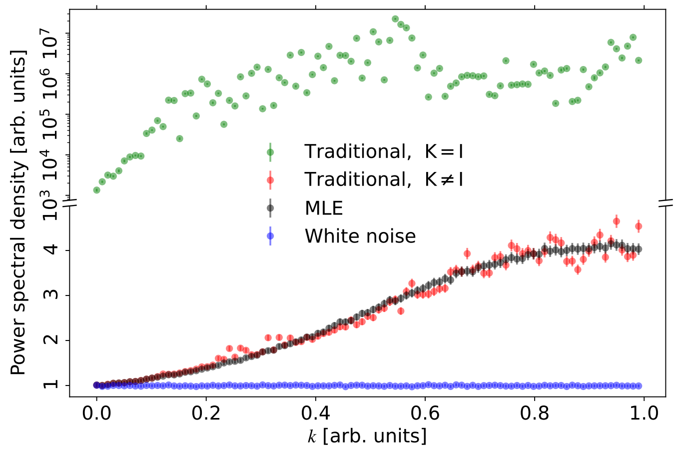


Fig. 2. Comparison of noise power spectra for noise-only 1D simulations for four different analysis configurations, as evaluated by the mean and standard deviation from 10 000 simulations. The y-axis is broken into linear and logarithmic portions.

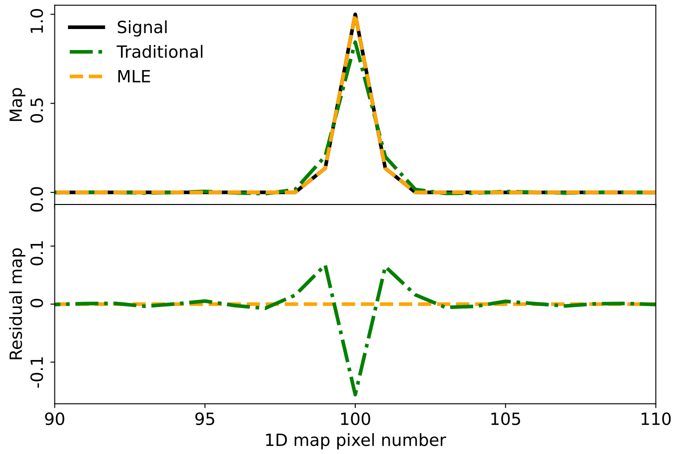


Fig. 4. Comparison of reconstructed 1D point source signals for $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ (dashed green) and $\hat{\boldsymbol{m}}_{\mathrm{MLE}}$. The top panel shows the full reconstructed signal amplitude, with the true input shown as a solid black line, and the bottom panel shows the difference between output and input signals.

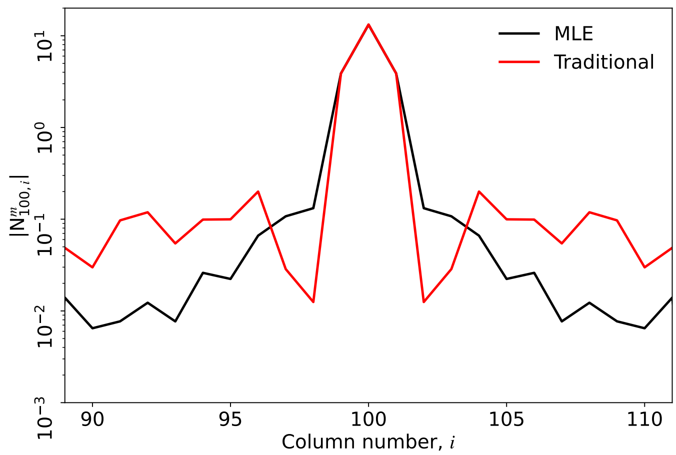


Fig. 3. Absolute value of a middle row map covariance matrix slice, $|\mathsf{N}^{m}_{100,i}| = |(\mathsf{P}^T\mathsf{N}^{-1}\mathsf{P})^{-1}_{100,i}|$, for both the 1D MLE (black) and binned and low-pass filtered solutions of the 1D toy model (red). These covariances have been found from the corresponding map-domain power spectra seen in Fig. 2.

Eqs. (4) and (5) for $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ and $\hat{\boldsymbol{m}}_{\mathrm{MLE}}$, respectively. Since the number of pixels is only 200, these solutions are very fast even with brute-force matrix inversion.

We consider two different cases for this 1D model. In the first case, we set $\boldsymbol{s} = 0$, with the goal of understanding the effects of the different deconvolution methods on the noise properties of the resulting maps. In the second case, we insert a single narrow Gaussian peak in the middle of the map, representing a point source in a typical CMB map, with the goal of understanding the relative impact of the two methods on a sharp signal.

Starting with the noise-only case, we simulate a total of 10 000 independent noise realizations, and use these to build up the post-solution noise covariance matrix explicitly. For each realization, we calculate the power spectrum defined as $P(k) = \langle |f(k)|^2 \rangle$, where $f(k)$ is the Fourier transform of $\hat{\boldsymbol{m}}$. In the case of $\hat{\boldsymbol{m}}_{\mathrm{trad}}$, we have to deconvolve the effective transfer function arising from K to obtain an unbiased estimate. This is found by simulating a large ensemble of random signal-only maps and taking the ensemble average of the ratio between the corresponding output and input spectra. To illustrate the adverse impact of unregularized high-frequency noise on $\hat{\boldsymbol{m}}_{\mathrm{trad}}$, we also include a case corresponding to $\mathsf{K} = \mathsf{I}$ for $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ in this demonstration.

The results from these calculations are summarized in Fig. 2. For reference, the blue curve shows the power spectrum of a white noise TOD directly binned into a map, without taking into account T; this illustrates the intrinsic noise level that is subsequently boosted by the bolometer transfer function deconvolution in the actual methods. Starting from the top, the green dots show the noise in $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ when not applying the regularization kernel K. The y-axis in the plot is broken into linear and logarithmic scaling. The high-frequency noise must be suppressed prior to mapmaking in some way or other to obtain meaningful results. Moving on to the realistic cases corresponding to $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ with K and $\hat{\boldsymbol{m}}_{\mathrm{MLE}}$ shown in red and black, respectively, we see that the two methods perform similarly in terms of total noise power.

However, even though the two methods perform similar in terms of absolute noise power, they still perform quite differently in terms of noise correlations. This is illustrated in Fig. 3, which shows a slice through the empirical correlation matrix evaluated as $\langle \hat{\boldsymbol{m}}_j \hat{\boldsymbol{m}}_j^T \rangle$ over the simulated ensemble, where j indicates simulation number. We see that the correlation values fall off by almost an order of magnitude lower for $\hat{\boldsymbol{m}}_{\mathrm{MLE}}$ compared to $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ at long distances; the low absolute values for $\hat{\boldsymbol{m}}_{\mathrm{trad}}$ at a few pixels separation is just a ringing artifact from the K filter.

Next, we consider the signal case with a Gaussian point source in the middle of the 1D map. We repeat the same procedure as outlined in the previous section for both techniques. However, since we are now interested in the effect on the signal, and the mapmaking equations are linear, we now omit the noise in the actual simulated TOD. The resulting products therefore correspond directly to ensemble-averaged quantities, and require no Monte Carlo simulation. The outputs from these calculations are summarized in Fig. 4. The top panel shows the input model as a solid black curve, and the reconstructed estimates are shown as dashed orange (for $\hat{\boldsymbol{m}}_{\mathrm{MLE}}$) and dashed green (for $\hat{\boldsymbol{m}}_{\mathrm{trad}}$) curves. The bottom panel shows the difference between output and input signals. The input signal is normalized to unity at the peak, so that the bottom panel can be interpreted as a fractional error. The MLE solution results in an unbiased estimate of the input signal, and any uncertainties in this solution are given by numerical round-off errors. In contrast, the traditional method results in residuals at the 10% level at the peak, with significant ringing extending to large distances.

4. Two-dimensional toy model: Impact on effective beam

In this section we apply the methods outlined in Sect. 2 to a two-dimensional case, with the goal of comparing the performance of the traditional and the MLE methods in terms of their impact on the effective beam. In this case, we construct an input sky signal that is mostly empty, except for one or more bright point sources depending on the test in question. In the first case we study a single point source located at the Ecliptic South Pole, and the input map is defined by setting the closest pixel to 100 in arbitrary units. The map is then smoothed with a symmetric Gaussian beam with the full width at half maximum (FWHM) set to 7.2 arcmin, similar to the 143 GHz *Planck* HFI effective beam (Planck Collaboration VII 2014).

Using this map, the sky signal TOD is created using the *Planck* 143-5 pointing matrix as $\boldsymbol{s} = \mathsf{P}\boldsymbol{m}_{\text{sky}}$. For the purposes of this experiment, which is designed to build intuition regarding the effect of a bolometer transfer function on a point source, we only use the first three months of the HFI survey. The measured TOD $\boldsymbol{d}$ is created through Eq. (1) by applying the bolometer transfer function and adding white noise. The white noise level is set to $\sigma_{\text{wn}} = 0.3125$ in arbitrary units, which corresponds to a signal-to-noise ratio of 320 at the peak of the point source. This value is chosen to represent the typical signal-to-noise ratio of planet observations reported by Planck Collaboration VII (2014). The maps for both the input sky signal $\boldsymbol{s}$ and the corresponding naively binned sky map $\hat{\boldsymbol{m}}_{\text{bin}}$ are shown in Fig. 5. In the first case, the point source appears azimuthally symmetric, while in the second case it is significantly deformed. Because the satellite scanning moved from right to left in this figure, the transfer function effectively drags the signal along the scanning path.

We now apply both the traditional and the MLE methods to these simulated data, each producing a map of the true sky signal. The results from these calculations are shown in Fig. 6 in terms of the difference between the reconstructed and the true input maps. Starting with the traditional method, we observe at least two noticeable residuals related to the transfer function. First, since the traditional method includes a regularization kernel K, the algorithm is unable to reconstruct the true input sky signal, and a quadrupolar residual aligned with the scanning path is present in the residual map. For an isotropic and random field, such as the CMB, the average effect of this can be accounted for by modifying the effective azimuthally symmetric beam response function b_ℓ, as for instance is done in the *Planck* analysis, but this method is clearly unable to reconstruct an optimal image of the true sky. In contrast, the integrated MLE solution shows no signs of scan-aligned effects, and the residual is consistent with white noise. Secondly, far away from the point source, the traditional method smooths the white noise more than the MLE method. Both of these effects correspond directly to what was found for the one-dimensional toy model in the previous section.

The magnitude of this effect depends on the detailed scanning path of the instrument, and hence changes with the position on the sky. We now aim to quantify the effective beam deformation produced by the bolometer transfer function for both methods in terms of the effective beam FWHM and ellipticity over the full sky. For these purposes a statistically meaningful number of point sources is required. Therefore, we create

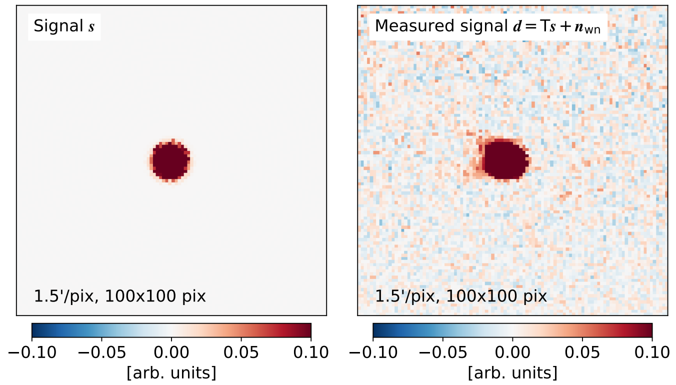


Fig. 5. Simulated point source signal $\boldsymbol{s}$ on the *left* panel and detector measured signal $\boldsymbol{d} = \mathsf{T}\boldsymbol{s} + \boldsymbol{n}$ on the *right* panel. The effect of the transfer function T can be observed in the smearing of the point source signal along the scanning path, resulting in a deformed beam.

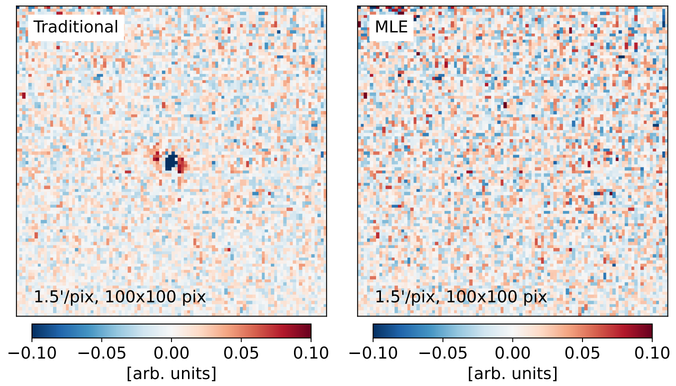


Fig. 6. Residual comparison between the traditional deconvolution method and MLE based deconvolution method. The *left panel* shows the residual between the signal deconvolved using the inverse of the transfer function operator T^{-1} in combination with filter $K(\omega)$ and the original point source signal $\boldsymbol{s}$ (*left panel* in Fig. 5). The *right panel* shows the residual between the signal deconvolved by solving Eq. (5) and the original point source signal.

a map with 12 288 point sources, each located at the center of a HEALPix $N_{\text{side}} = 32$ map. The analysis itself is performed with $N_{\text{side}} = 2048$, corresponding to a pixel size of 1′.7. Then we apply exactly the same process as earlier in this section in terms of transfer function operator T, white noise, and map-making methods.

For each point source and both methods, we measure the effective FWHM and ellipticity by fitting a two-dimensional Gaussian following the steps similar to Fosalba et al. (2002). We define the ellipticity parameter $\varepsilon = \sigma_{\text{long}}/\sigma_{\text{short}}$ as the ratio between the long and short axes of the ellipse. An azimuthally symmetric object corresponds to $\varepsilon = 1$, while $\varepsilon > 1$ corresponds to a deformed beam. In terms of polar coordinates (ρ, ϕ), the actual function fitted to each two-dimensional object is

$$z(\rho, \phi) = A \cdot \exp\left[-\frac{\rho^2}{2\sigma_{\text{short}}^2}\left(1 - \chi \cdot \cos(\phi - \alpha)^2\right)\right], \tag{7}$$

where σ_{short} is the width of the short axis of the ellipse, $\chi \equiv 1 - 1/\varepsilon^2$, and α is the rotation angle to align coordinate axes with the ellipse axes. The effective FWHM is defined as $\sqrt{8\ln 2}$ times the average between the long and short axes widths, where long axis width can be found from the relation for ellipticity $\sigma_{\text{long}} = \varepsilon \cdot \sigma_{\text{short}}$.

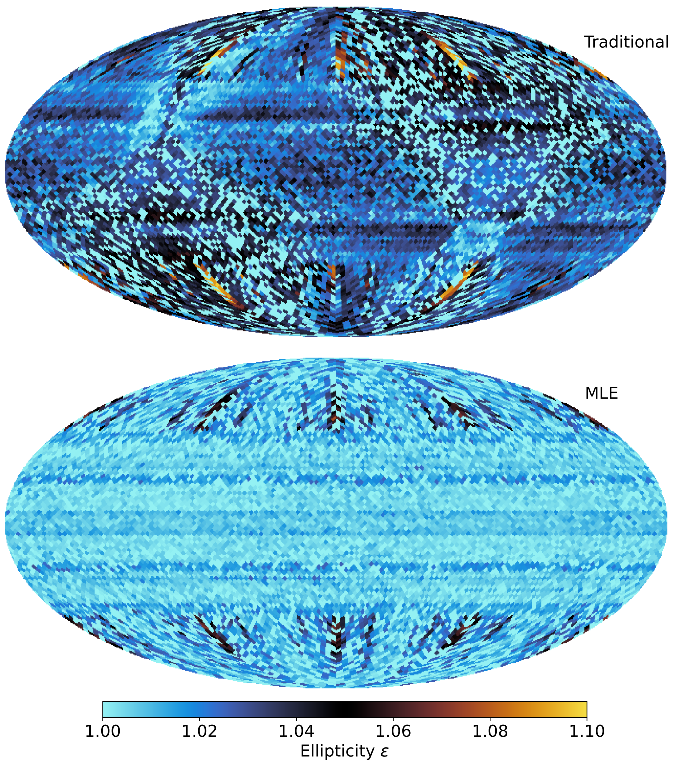


Fig. 7. Distribution of ellipticity $\varepsilon = \sigma_{\text{long}}/\sigma_{\text{short}}$ over the sky. The ellipticity was measured as a parameter in a Gaussian fit in Eq. (7). The *upper panel* shows the ellipticity of the beams deconvolved using traditional method, and the *lower panel* – deconvolved with MLE method.

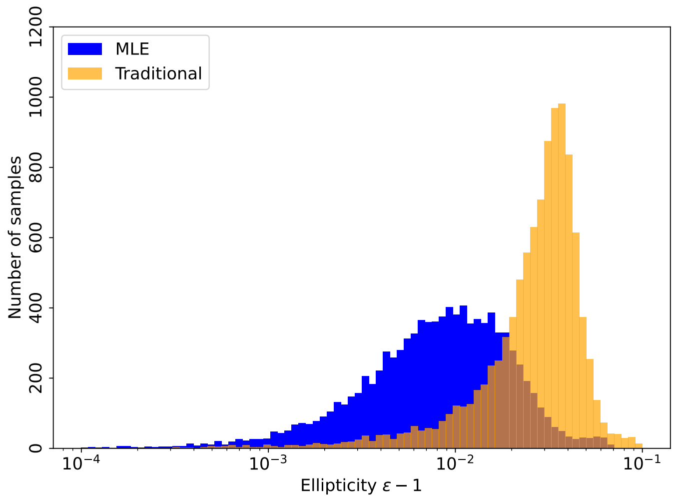


Fig. 8. Distribution of effective ellipticities, $\varepsilon - 1$, for the traditional (orange histogram) and MLE mapmaking (blue histogram) algorithms.

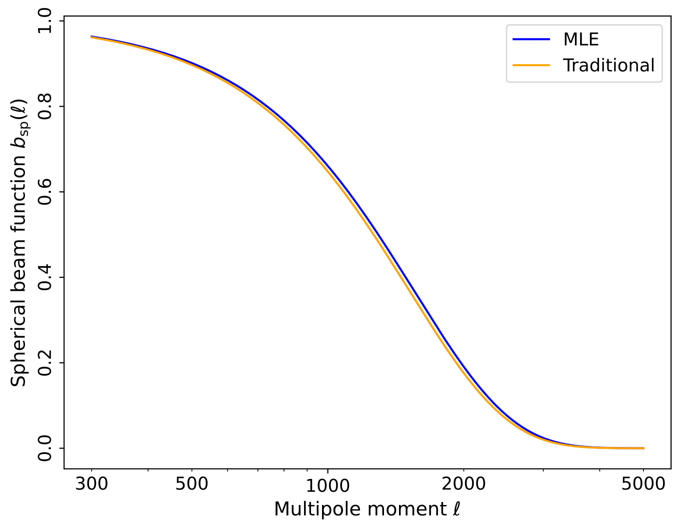


Fig. 9. Spherical beam function for the two deconvolution methods. These functions are calculated based on the average FWHM for each method. The blue line shows the spherical beam function for the maximum likelihood deconvolution method, while the orange line shows *Planck* method.

The function defined in Eq. (7) assumes that the coordinate center is located at the peak, and this is not necessarily true after beam convolution. We therefore run an additional fit for the center pixel coordinates from the $N_{\text{side}} = 32$ map to the center coordinates of these beams on the $N_{\text{side}} = 2048$ map. We can assume local space around a given beam to be Euclidean. Then Eq. (7) follows from a regular two-dimensional normal distribution in Cartesian coordinates after the polar coordinate transformation. In order to align the axes of the ellipse with the smearing effect produced by the deconvolution, a rotational angle α is introduced into the fitting function.

In Fig. 7 we show the distribution of ε over the full sky for both mapmaking methods, and we notice that the traditional method results in a noticeably higher ellipticity across the sky compared to the MLE method proposed in this paper, and it has a much stronger imprint of the *Planck* scanning strategy. In contrast, the distribution seen for the MLE method is defined primarily by the underlying HEALPix grid, which is unavoidable given the choice of pixelization.

Figure 8 shows the same information in terms of histograms of $\varepsilon - 1$. The corresponding means and standard deviations for the two distributions are $\varepsilon_{\text{trad}} - 1 = 0.025 \pm 0.014$ and $\varepsilon_{\text{MLE}} - 1 = 0.009 \pm 0.010$. The mean ellipticity of the MLE method is thus 65% smaller than for the traditional method. The values found for the traditional method are close to those reported by Planck Collaboration VII (2014) for the 143 GHz channel.

Performing a similar comparison for the effective FWHM, we find that the MLE method results in a 2.3% lower value than the traditional method. The net impact of this difference in terms of effective beam transfer functions, b_ℓ, is shown in Fig. 9, where

$$b_{\text{sp}}(\ell) = \exp\left(-\frac{\ell(\ell+1)\text{FWHM}}{16\ln 2}\right). \tag{8}$$

At $\ell = 2500$, the ratio between these two functions is 1.14, while at $\ell = 4000$ it is 1.38.

5. CMB simulation

Finally, we consider a semi-realistic CMB-plus-noise case. In this case, the sky signal $\boldsymbol{s}$ is generated as a Gaussian random realization based on a best-fit ΛCDM temperature power spectrum computed with CAMB (Lewis et al. 2000; Howlett et al. 2012) and adopting best-fit parameters from Planck Collaboration V (2020). The simulated TOD is then generated by observing this map with the full-mission *Planck* 143-5 scanning path, and applying the corresponding bolometer transfer function T. Finally, we add white noise with $\sigma = 200\,\mu\text{K}$ per sample. This

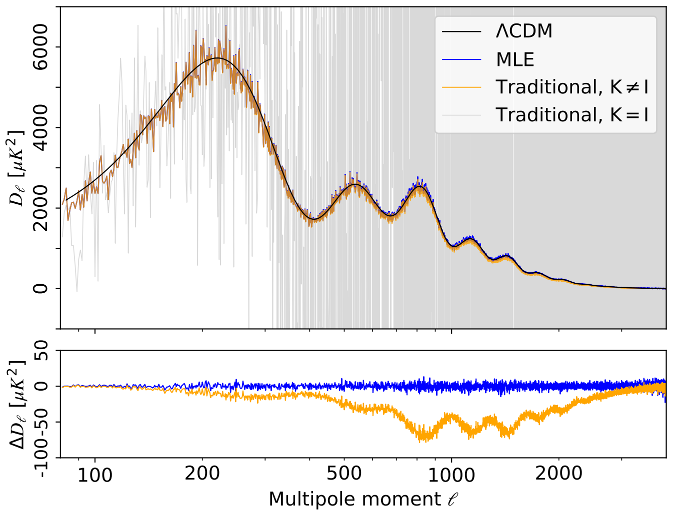


Fig. 10. *Upper panel*: power spectra D_ℓ of the deconvolved CMB maps. The original signal was simulated based on the ΛCDM spectrum shown with the black solid line. The blue line shows the power spectrum obtained from the maximum likelihood based deconvolution. The orange line shows the power spectrum of the map obtained from the traditional deconvolution process $\mathsf{K}\mathsf{T}^{-1}\boldsymbol{d}$. The grey line shows the power spectrum of the map obtained from the unfiltered ($\mathsf{K} = \mathsf{I}$) traditional deconvolution process $\mathsf{T}^{-1}\boldsymbol{d}$. All spectra are calculated as cross power spectra using two noise realisations $\boldsymbol{n}_{\mathrm{wn}}^{1,2}$, to exclude the white noise power spectrum. *Lower panel*: difference between the deconvolved power spectra and input ΛCDM power spectrum. The colors represent the difference for the respective deconvolution method in the *upper* panel.

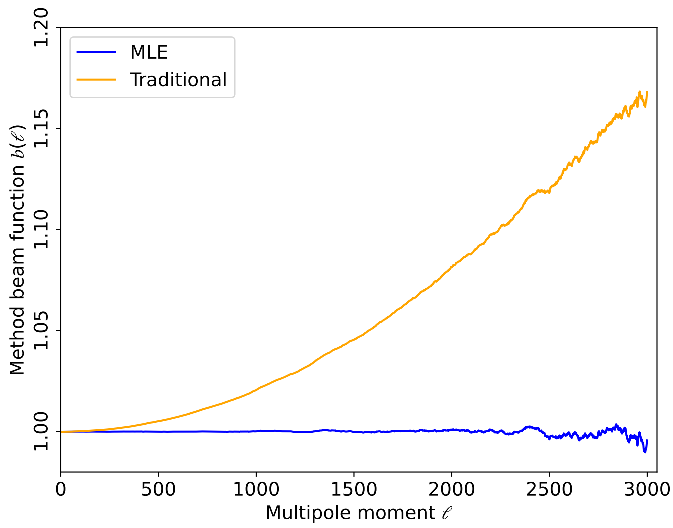


Fig. 11. Effective beam function introduced by the deconvolution method. This function is found as the square of a ratio between the simulated ΛCDM power spectrum and the power spectrum calculated for the respective deconvolution method. The maximum likelihood based method produces a beam function equal to one, as shown by the blue line. The traditional deconvolution method results in the beam function increasing with the multipole moment ℓ and is illustrated with the orange line.

value corresponds to coadding all 143 GHz bolometers into one, such that our final simulation has similar sensitivity as the true 143 GHz frequency map, but the data volume of only a single detector. To allow for the calculation of cross-power spectra, we split the data into two halves, and process each half independently.

We now apply the same three map-making methods to this TOD simulation as shown in Fig. 2 for the one-dimensional case, namely the traditional method (with and without a regularization kernel) and the new MLE method. The results from this calculation are summarized in Fig. 10 in terms of cross-angular power spectra $D_\ell = C_\ell \ell(\ell + 1)/2\pi$.

The traditional method without a low-pass filter K (gray) produces results that are extremely noisy even after taking the cross-power spectrum, mirroring the 1D case shown in Fig. 2. Secondly, we see that both the traditional and MLE methods reproduce the original ΛCDM power spectrum at low multipoles, $\ell \lesssim 100$. However, at higher multipoles the traditional method is noticeably lower. This deviation is caused by the additional K smoothing operator, which has not been deconvolved in this plot.

In order to correct for this bias, one has to include the effect of K into the effective beam profile, as for instance done by Planck Collaboration VII (2014). To measure the total beam profile, we simply calculate the square root of the ratio between output and input power spectra. The results are shown in Fig. 11. Here we see that the MLE method produces an effective beam profile that is very close to unity for almost all multipoles. The small deviations seen at higher ℓ are due to instrumental and numerical noise. On the other hand, the traditional method results in a ratio that monotonically increases with ℓ.

6. Computational expense

The code used in this this paper was written in Python, and relies on utilities provided by the `scipy` (Virtanen et al. 2020) and `numpy` (Harris et al. 2020) libraries. The majority of the runtime is spent on fast Fourier transforms (FFT), which are performed with a compiled C++ code under the hood. However, the emphasis in this paper has been the fundamental algebraic solution, and not code optimization, and the runtimes quoted in the following can very likely be improved by a significant factor in a future production implementation.

Overall, the total cost for the MLE solution is given by the product of the cost for a single CG iteration and the total number of CG iterations. The number of CG iterations depends in turn on the noise level of the data, and the runtimes below are given for the full-sky and full-mission *Planck* HFI 143 GHz case. For a noise level of $\sigma = 200\,\mu$K, the algorithm requires 29 iterations to converge using the preconditioner from Eq. (6), with a convergence criterion defined by a relative error of $\delta_{\mathrm{new}}/\delta_0 = 10^{-10}$. Each iteration costs 32.9 CPU-hrs, out of which 19.2 CPU-hrs are used on parallel FFT calculations and application of the transfer function $T(\omega)$ within the $\mathsf{T}^T\mathsf{N}^{-1}\mathsf{T}$ operator. In order to parallelise and speed up the Fourier transformations efficiently, the TODs are divided into overlapping segments of length 2^{19}, similar to Planck Collaboration VI (2014). The run required a total of 563 CPU-hrs, or 14.9 wall-hours when parallelized over 64 cores. We note here, that only a part of that time is spent in parallel regime, while the other part is not parallelized. Ultimately, this algorithm is intended to be integrated in the end-to-end CMB Gibbs sampler `Commander` (Galloway et al. 2023). For comparison, the cost for a full Gibbs sample for all three *Planck* LFI channels was 169 CPU-hrs (BeyondPlanck Collaboration I 2023; Galloway et al. 2023; Basyrov et al. 2023). Assuming no further optimizations, full analysis of all *Planck* HFI channels with this algorithm will increase the total runtime by more than an order of magnitude.

Full optimization will happen in the future Fortran implementation, and for now we simply conclude that this algorithm is indeed feasible, even though it is computationally expensive.

7. Conclusions

In this work we have proposed a new method for accounting for the finite bolometer transfer function in modern bolometer-based CMB experiments. This method integrates the bolometer transfer function directly into the classical maximum-likelihood mapmaking equation, which then is solved with a conjugate gradient method. This method is the optimal solution to the full deconvolution problem, and provides an unbiased signal estimate with proper noise weighting and correlations.

We have compared this method to the traditional two-step procedure used by most bolometer-based experiments to date, in which an estimate of the transfer function is deconvolved from the TOD prior to mapmaking. For slow detectors with a long bolometer time constant compared to the sampling rate, the deconvolution procedure boosts the white noise at high temporal frequencies, and this is usually regularized explicitly by an explicit and additional smoothing kernel. We have shown that the resulting map estimate contains both significant scan-aligned residuals and larger noise correlations than the optimal method discussed in this paper. Calculating the effective ellipticity and FWHM of the beams resulting from the two deconvolution methods, we find that the ellipticity is 64% lower for a *Planck* 143 GHz-based simulation with the MLE method than for the traditional method, and the FWHM is 2.3% smaller. Another notable advantage of the optimal method is that it does not require an additional power spectrum level deconvolution kernel (because $\mathsf{K} = \mathsf{I}$ in this case), which should result in significantly lower beam estimation uncertainties when integrated into a full pipeline.

Based on these findings, we conclude that the new method is algebraically preferable over the traditional method. At the same time, the computational cost is also correspondingly higher, with a total runtime of many hundreds of CPU-hours for a typical *Planck* HFI case. However, this cost will likely be decreased significantly both through better code optimization and algebraic improvements, for instance by implementing a better CG preconditioner. This is left for future work. In addition, the current implementation already results in runtimes that are fully feasible for modern computers. We anticipate that this method will allow for better data extraction both in reanalyses of the previous CMB experiments, such as *Planck* HFI, and the analysis of the upcoming ones, such as Simons Observatory, LiteBIRD, and CMB-S4.

Acknowledgements. The current work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement numbers 819478 (ERC; COSMOGLOBE), 772253 (ERC; BITS2COSMOLOGY), and 101007633 (Marie Skłodowska-Curie, CMB-INFLATE). In addition, the collaboration acknowledges support from RCN (Norway; grant no. 274990). Some of the results in this paper have been derived using the `healpy` and `HEALPix` (`http://healpix.sf.net`) packages (Górski et al. 2005; Zonca et al. 2019).

References

Basyrov, A., Suur-Uski, A.-S., Colombo, L. P. L., et al. 2023, A&A, 675, A10
Bennett, C. L., Banday, A. J., Gorski, K. M., et al. 1996, ApJ, 464, L1
Bersanelli, M., Mandolesi, N., Butler, R. C., et al. 2010, A&A, 520, A4
BeyondPlanck Collaboration I. 2023, A&A, 675, A1
Fosalba, P., Doré, O., & Bouchet, F. R. 2002, Phys. Rev. D, 65, 063003
Galloway, M., Andersen, K. J., Aurlien, R., et al. 2023, A&A, 675, A3
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, JCAP, 1204, 027
Ihle, H. T., Bersanelli, M., Franceschet, C., et al. 2023, A&A, 675, A6
Keihänen, E., & Reinecke, M. 2012, A&A, 548, A110
Lamarre, J., Puget, J., Ade, P. A. R., et al. 2010, A&A, 520, A9
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Planck Collaboration I. 2014, A&A, 571, A1
Planck Collaboration VI. 2014, A&A, 571, A6
Planck Collaboration VII. 2014, A&A, 571, A7
Planck Collaboration I. 2016, A&A, 594, A1
Planck Collaboration VII. 2016, A&A, 594, A7
Planck Collaboration I. 2020, A&A, 641, A1
Planck Collaboration V. 2020, A&A, 641, A5
Shewchuk, J. R. 1994, An Introduction to the Conjugate GradientMethod Without the Agonizing Pain, Edition $1\frac{1}{4}$, `http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf`
Stevens, J. R., Cothard, N. F., Vavagiakis, E. M., et al. 2020, J. Low Temp. Phys., 199, 672
Tegmark, M. 1997, ApJ, 480, L87
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261
Zhao, Y., Allen, C., Amiri, M., et al. 2008, in Millimeter and Submillimeter Detectors and Instrumentation for Astronomy IV, eds. W. D. Duncan, W. S. Holland, S. Withington, & J. Zmuidzinas, SPIE Conf. Ser., 7020, 70200O
Zonca, A., Singer, L. P., Lenz, D., et al. 2019, J. Open Source Software, 4, 1298