Deliverable 1.3: Final Gibbs/posterior products

Authors Hans Kristian Eriksen

Date November 19th, 2020

Work Package WP1 – Gibbs sampling

DocId [xxx-xxx-xxx]

Revision History

Version	Authors	Date	Changes
1.0	Hans Kristian Eriksen	November 19th, 2020	Initial Version

Contents

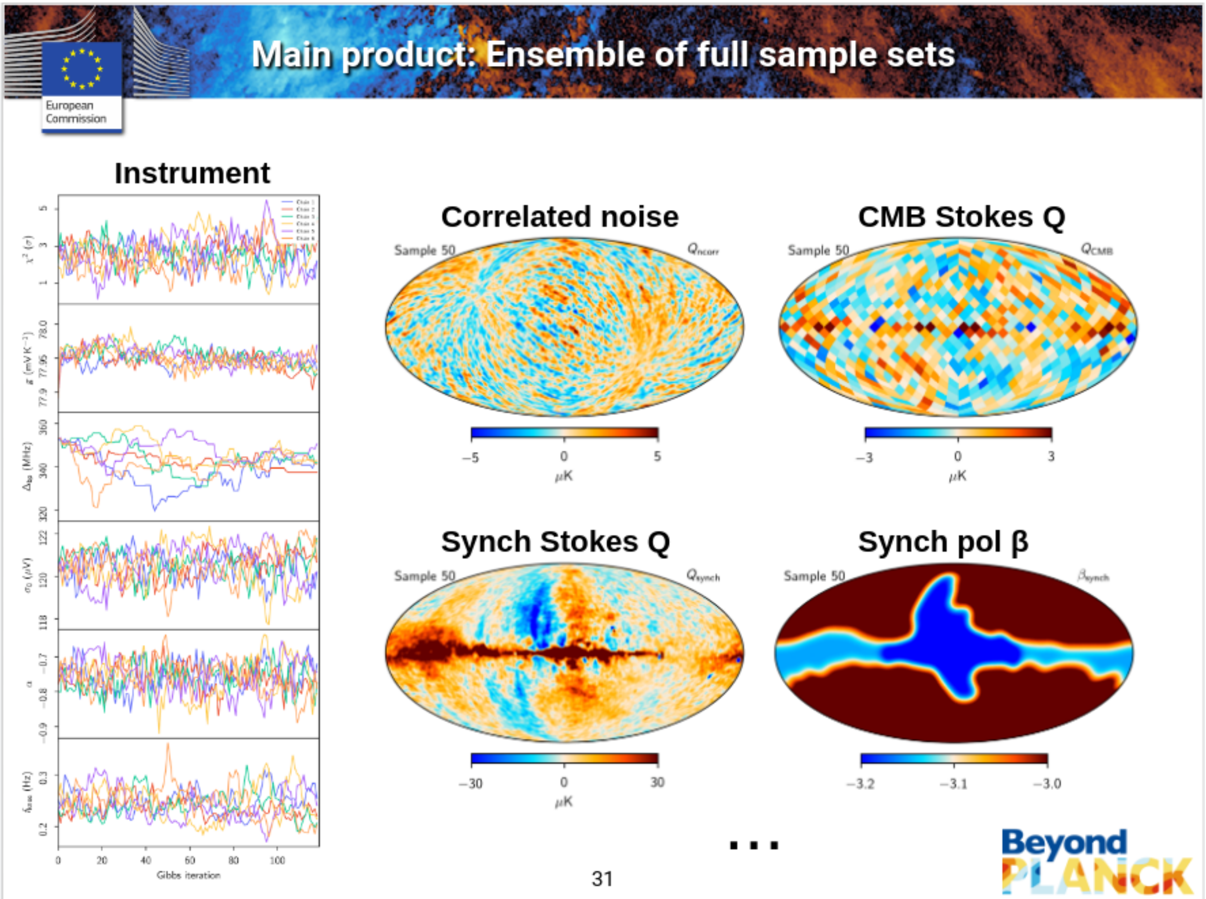

Figure 1: Slide from BeyondPlanck overview talk at the release conference from November 18-20. This shows a snapshot of the final Gibbs samples/posterior distribution of the final data products.

1 Overview

The BeyondPlanck collaboration proposed more than three years ago to implement the first end-to-end Gibbs sampler for CMB analysis experiments, and apply this to the Planck LFI data set. This has now been successfully completed, and the results were presented to the public in an online release conference on November 18-20, 2020, attended by more than 250 scientists from 26 countries in 6 continents. The event was a great success, and both the BeyondPlanck collaboration members and the community in general has learned much about how one can organize online conferences safely and with low cost even during a pandemic. Figure 1 shows one of the slides shown at this conference, which summarizes some selected parameters from the BeyondPlanck posterior distribution.

The BeyondPlanck data products are available through the project webpage (https://beyondplanck.science -> Products). The main product set are three sets of so-called “chain files”, which each contain sets of Gibbs samples. The three sets correspond to 1) the main analysis, including all time-ordered data, astrophysical parameter and frequency sky maps; 2) re-sampled high-resolution CMB temperature samples; and 3) re-sampled low-resolution CMB temperature+polarization samples. The main product is thus the first set, while the other two cases explore the CMB parameter volume with denser sampling, in order to derive more robust cosmological parameter estimates. Specifically, while the main chains only contain 200 samples each (for a total of 1200 complete samples), the low-l chains contain more than 7000 samples each (for a total of about 45,000 samples). All results are described in a suite of scientific papers available from the project webpage (https://beyondplanck.science -> Publications)

When we started this work, it was not clear how many samples would actually be sufficient in order to obtain robust convergence with respect to important parameters such as the optical depth of reionization, although intuition based on the number of relevant multipoles suggested that about 1000 would be appropriate. As discussed in the current paper suite, this turned out to be a good guess, and we have indeed reached acceptable convergence. However, the buffer is not large, as the final values did not settle down until after about 700 samples, which means that it is not currently possible to split the data set in two, and cross-check that one obtains the same answer with two independent sample sets. Future analyses should therefore ideally aim for 1500-2000 samples, rather than 1000. This is an important lesson learned from the current analysis.

The contents and format of the chains files are specified in the main documentation, which is available at http://docs.beyondplanck.science. All scientific results are described by BeyondPlanck Collaboration (2020) and references therein.

2 Production schedule and delivery

According to the original schedule, the final BeyondPlanck products were supposed to be completed three monts prior to the end of the project, i.e., on August 31 2020. Indeed, the very first complete set of BeyondPlanck products were completed before this time, after having run through the summer vacation weeks. However, when we inspected the results, it was clear that those results are sub-optimal and not suitable for official publication. For instance, the correlated noise parameters showed clear signs of algorithmic bias, and the CMB TT spectrum was obviously biased at intermediate angular scales.

In the beginning of September, we therefore made the radical but absolutely necessary decision that we had to make several important adjustments to the algorithm, and then restart the production. The two biggest changes were 1) a full implementation of an exact Gaussian constrained realization (Wiener filter) solver for correlated noise (the summer version used an approximate polynomial gap filling, which is commonly used in the CMB field; this turned out to be woefully inadequate for our purposes) and 2) active spatial priors for free-free and AME emission (which solved the CMB TT bias problem).

Of course, given that the raw production time is 3-4 weeks, this put the entire collaboration on an extremely tight schedule with respect to the final release conference – but it worked, and through many long hours the collaboration managed to pull it all together in time for the release. A lesson learned for future projects of this type is to allow sufficient time for multiple reprocessings of the full data set; two is an absolute bare minimum, and three would be better. Each processing obviously carries an additional cost in terms of computing and work power, but there are always surprises, and it takes time to understand and fix these.

Product delivery to the Planck Legacy Archive (PLA) has already been agreed with Dr. Jan Tauber, and will take place as soon as the overview paper has been accepted for publication. According to PLA policies, all products featured on their webpage must have been peer-reviewed before publication, and this is understood by all parties. Given the complexity of the overview paper, a likely PLA delivery date is March 2021. However, all products are of course already now available through the project homepage, and so this delay does not represent any significant problem for the community at large.

3 Outstanding problems

Overall, the current BeyondPlanck products truly establishes a new state-of-the-art for understanding the microwave sky between 30 and 70GHz. The new LFI maps are significantly cleaner than all previous incarnations, both from the LFI Data Processing Center and from NPIPE, and it is now finally possible to analyze both Planck LFI and WMAP polarization data jointly, and obtain a physically plausible sky model.

At the same time, as reported in the BeyondPlanck publications, there are two specific outstanding problems that has been identified for the first time in the BP processing, but not yet solved. These are 1) gain-correlated stripes at 44 GHz (Gjerløw et al. 2020), and 2) excess correlated noise in both 30 and 44 GHz (Ihle et al. 2020). A main goal for future post-BeyondPlanck processing is to understand, model and mitigate these issues. Indeed, useful initiatives along these lines were already made during the BeyondPlanck conference by Dr. Bob Watson, who was deeply involved in the original LFI analysis of this channel. This is a useful example of how the BeyondPlanck work has already generated new activity within the larger community.

4 References

BeyondPlanck I. Global Bayesian analysis of the Planck Low Frequency Instrument data, BeyondPlanck Collaboration, 2020, A&A, submitted, [2011.05609]

BeyondPlanck VI. Noise characterization and modelling, Ihle, H. T. et al. 2020, A&A, submittted [2011.06650]

BeyondPlanck VII. Bayesian estimation of gain and absolute calibration for CMB experiments, Gjerløw, E. et al. 2020, A&A, submitted [2011.08082]