

# COSMOGLOBE I. Improved Wilkinson Microwave Anisotropy Probe frequency maps through Bayesian end-to-end analysis

D. J. Watts<sup>1\*</sup>, A. Basyrov<sup>1</sup>, J. R. Eskilt<sup>1,2</sup>, M. Galloway<sup>1</sup>, L. T. Hergt<sup>3</sup>, D. Herman<sup>1</sup>, H. T. Ihle<sup>1</sup>, S. Paradiso<sup>4</sup>, F. Rahman<sup>5</sup>, H. Thommesen<sup>1</sup>, M. Bersanelli<sup>6</sup>, L. A. Bianchi<sup>6</sup>, M. Brilenkov<sup>1</sup>, L. P. L. Colombo<sup>6</sup>, H. K. Eriksen<sup>1</sup>, C. Franceschet<sup>6</sup>, U. Fuskeland<sup>1</sup>, E. Gjerløw<sup>1</sup>, B. Hensley<sup>8</sup>, G. A. Hoerning<sup>7</sup>, K. Lee<sup>1</sup>, J. G. S. Lunde<sup>1</sup>, A. Marins<sup>7,9</sup>, S. K. Nerval<sup>10,11</sup>, S. K. Patel<sup>12</sup>, M. Regnier<sup>13</sup>, M. San<sup>1</sup>, S. Sanyal<sup>12</sup>, N.-O. Stutzer<sup>1</sup>, A. Verma<sup>12</sup>, I. K. Wehus<sup>1</sup>, and Y. Zhou<sup>14</sup>

<sup>1</sup> Institute of Theoretical Astrophysics, University of Oslo, Blindern, Oslo, Norway

<sup>2</sup> Imperial Centre for Inference and Cosmology, Department of Physics, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, United Kingdom

<sup>3</sup> Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver BC, V6T1Z1, Canada

<sup>4</sup> Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada

<sup>5</sup> Indian Institute of Astrophysics, Koramangala II Block, Bangalore, 560034, India

<sup>6</sup> Dipartimento di Fisica, Università degli Studi di Milano, Via Celoria, 16, Milano, Italy

<sup>7</sup> Instituto de Física, Universidade de São Paulo - C.P. 66318, CEP: 05315-970, São Paulo, Brazil

<sup>8</sup> Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08540

<sup>9</sup> Department of Astronomy, University of Science and Technology of China, Hefei, China

<sup>10</sup> David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

<sup>11</sup> Dunlap Institute for Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

<sup>12</sup> Department of Physics, Indian Institute of Technology (BHU), Varanasi - 221005, India

<sup>13</sup> Laboratoire Astroparticule et Cosmologie (APC), Université Paris-Cité, Paris, France

<sup>14</sup> Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA

March 8, 2023

## ABSTRACT

We present the first joint analysis of *WMAP* and *Planck* LFI time-ordered data, processed within the Bayesian end-to-end Commander framework. This framework builds directly on a similar analysis of the LFI measurements by the BEYONDPLANCK collaboration, and approaches the CMB analysis challenge through Gibbs sampling of a global posterior distribution. The computational cost of producing one complete *WMAP*+LFI Gibbs sample is 812 CPU-hr, including calibration, mapmaking, and component separation, of which 603 CPU-hr is spent on *WMAP* low-level processing; this demonstrates that end-to-end Bayesian analysis of the *WMAP* data is computationally feasible. We find that our *WMAP* posterior mean temperature sky maps are largely consistent with the official maps, and the resulting CMB power spectrum is in excellent agreement with previous results. The most notable difference is a slightly lower CMB quadrupole amplitude of  $\sigma_2 = 132 \pm 71 \mu\text{K}^2$ , as compared to  $\sigma_2 = 229 \pm 97 \mu\text{K}^2$  in the BEYONDPLANCK analysis. In contrast, our *WMAP* polarization maps differ more notably from the official results, and in general they exhibit lower large-scale residuals, most likely attributable to a better constrained gain and transmission imbalance model; it is particularly noteworthy that our *W*-band sky maps appear statistically consistent with the *V*-band maps. For the first time, *WMAP*-minus-LFI frequency map differences appear visually consistent with instrumental noise over most of the sky. Still, we identify three specific issues that require additional work, namely 1) low-level noise modeling, 2) quadrupole residuals in the *V*- and *W*-band temperature maps at the  $2 \mu\text{K}$  level; and 3) a strong degeneracy between the absolute *K*-band calibration and the dipole of the anomalous microwave emission component. Nevertheless, we believe that the reprocessed *WMAP* maps presented here are significantly cleaner in terms of systematic uncertainties than the official *WMAP* maps. Both sky maps and the associated code are made publicly available through the COSMOGLOBE web page.

**Key words.** ISM: general – Cosmology: observations, polarization, cosmic microwave background, diffuse radiation – Galaxy: general

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>	2.4	Sky model . . . . .	6
<b>2</b>	<b>End-to-end Bayesian CMB analysis</b>	<b>3</b>	2.5	Priors and poorly measured modes . . . . .	6
2.1	LFI and BEYONDPLANCK . . . . .	3	2.6	Posterior distribution and Gibbs sampling . . . . .	8
2.2	Official <i>WMAP</i> instrument model and analysis pipeline . . . . .	3	2.7	Sampling algorithms . . . . .	8
2.3	COSMOGLOBE instrument model . . . . .	4	2.7.1	Review of sampling algorithms . . . . .	8
			2.7.2	Differential mapmaking . . . . .	9
			2.7.3	Transmission imbalance estimation . . . . .	10
			2.7.4	Baseline sampling . . . . .	10
			<b>3</b>	<b>Data and data processing</b>	10
			3.1	Publicly available <i>WMAP</i> products . . . . .	11

\* Corresponding author: D. J. Watts; [duncanwa@astro.uio.no](mailto:duncanwa@astro.uio.no)

3.2	TOD preprocessing and data selection . . . . .	11
3.3	Computational resources and future plans . . . . .	12
<b>4</b>	<b>Instrumental parameters</b>	<b>12</b>
4.1	Markov chains, correlations and posterior mean statistics . . . . .	13
4.2	Gain and baselines . . . . .	14
4.3	Transmission imbalance . . . . .	15
4.4	Instrumental noise . . . . .	16
4.5	Instrumental corrections in map domain . . . . .	19
4.6	Instrumental uncertainties in power spectrum domain . . . . .	21
<b>5</b>	<b>Frequency maps</b>	<b>21</b>
5.1	Map summary statistics . . . . .	21
5.2	Comparison with 9-year <i>WMAP</i> maps . . . . .	28
5.3	Consistency within <i>WMAP</i> channels . . . . .	30
5.4	Consistency between <i>WMAP</i> and LFI . . . . .	31
<b>6</b>	<b>Preliminary astrophysical results</b>	<b>31</b>
6.1	CMB results . . . . .	32
6.1.1	Solar dipole . . . . .	32
6.1.2	Angular temperature power spectrum . . . . .	33
6.1.3	Low- $\ell$ anomalies . . . . .	34
6.2	Galactic foregrounds . . . . .	36
6.2.1	Free-free . . . . .	36
6.2.2	Anomalous microwave emission . . . . .	36
6.2.3	Synchrotron . . . . .	37
6.2.4	$\chi^2$ and map space residuals . . . . .	38
6.3	<i>WMAP</i> -versus-LFI signal-to-noise ratio comparison . . . . .	38
<b>7</b>	<b>Outstanding issues</b>	<b>38</b>
7.1	Noise modeling . . . . .	38
7.2	V- and W-band quadrupole residual . . . . .	38
7.3	Degeneracy between K-band calibration and AME dipole . . . . .	38
7.4	Other minor effects . . . . .	41
7.4.1	Time-variable bandpass modeling . . . . .	41
7.4.2	Polarized sidelobe modeling . . . . .	41
<b>8</b>	<b>Conclusions</b>	<b>41</b>
<b>A</b>	<b>Survey of instrumental parameters</b>	<b>43</b>
A.1	Gain, baselines, noise and $\chi^2$ . . . . .	43
A.2	Transmission imbalance . . . . .	43
<b>B</b>	<b><i>WMAP</i> frequency map survey</b>	<b>43</b>
<b>C</b>	<b>Comparison with BEYONDPLANCK LFI results</b>	<b>43</b>

## 1. Introduction

The discovery of the cosmic microwave background (CMB) by Penzias & Wilson (1965) marked a paradigm shift in the field of cosmology, providing direct evidence that the Universe was once much hotter than it is today, effectively ruling out the steady-state theory of the universe (Dicke et al. 1965). This discovery spurred a series of ground-breaking cosmological experiments, including the Nobel Prize-winning measurements by COBE-FIRAS that confirmed the blackbody nature of the CMB (Mather et al. 1994) and COBE-DMR that measured temperature variations from the primordial gravitational field (Smoot et al. 1992).

The NASA-funded *Wilkinson Microwave Anisotropy Probe* (*WMAP*; Bennett et al. 2003a) mission was launched a decade after COBE-DMR, and mapped the microwave sky with 45 times higher sensitivity and 33 times higher angular resolution, and thereby revolutionizing our understanding of early universe physics (Bennett et al. 2003a). As quantified by Bennett et al. (2013), the permissible parameter space volume for a standard  $\Lambda$ CDM model was decreased by a factor of 68,000 by *WMAP*, and the best pre-*WMAP* determination of the age of the universe was  $t_0 < 14$  Gyr from Boomerang (Lange et al. 2001), with best-fit values of 9–11 Gyr; the latter values in apparent contradiction with direct measurements of the oldest globular clusters (Hu et al. 2001).

The ESA-led *Planck* satellite (Planck Collaboration I 2020) was developed concurrently with *WMAP*, and their operation lifetimes briefly overlapped, with *Planck* observing from 2009–2013 and *WMAP* from 2001–2011. *Planck*'s stated goal was to fully characterize the primary CMB temperature fluctuations from recombination, as well as to characterize the polarized microwave sky on large angular scales. Overall, *Planck*'s raw CMB sensitivity was an order of magnitude higher than *WMAP*'s, and its angular resolution three times higher. Today, *Planck* represents the state-of-the-art in terms of full-sky microwave sky measurements.

*Planck* comprised two independent experiments, namely the Low Frequency Instrument (LFI; Planck Collaboration II 2020) and High Frequency Instrument (HFI; Planck Collaboration III 2020), respectively. The LFI detectors were based on HEMT (high electron mobility transistor) amplifiers, spanning three frequency channels between 30 and 70 GHz, while the HFI detectors were based on TES (transition edge sensitive) bolometers, and spanned six frequency channels between 100 and 857 GHz. For comparison, *WMAP* was also HEMT-based, with comparable sensitivity to LFI alone, and spanned five frequencies between 23 and 94 GHz. At the same time, the two experiments implemented very different scanning strategies, and as a result they are highly complementary and synergistic; together they provide a clearer view of the low-frequency microwave sky than either can alone.

Towards the end of the *Planck* analysis phase it became clear that the interplay between instrument calibration and astrophysical component separation represented a main limiting factor in terms of systematic effects for high signal-to-noise measurements (Planck Collaboration II 2020). Specifically, in order to calibrate the instrument to sufficient precision, it became clear that it was necessary to know the true sky to a comparably high precision – but to know the sky, it was also necessary to know the instrumental calibration. The data analysis is thus fundamentally circular and global in nature. The final official *Planck* LFI analysis performed four complete iterations between calibration and component separation (Planck Collaboration II 2020), aiming to probe this degeneracy. However, it was clearly recognized that this was not sufficient to reach full convergence, and this sub-optimality led to the BEYONDPLANCK project (BeyondPlanck 2022), which aimed to perform thousands of complete analysis cycles, as opposed to just a handful. This framework was implemented using the Commander3 (Galloway et al. 2022a) code, a CMB Gibbs sampler that performs integrated high-level and low-level parameter estimation in a single integrated framework. This analysis demonstrated the feasibility of a full end-to-end Gibbs sampling analysis in the CMB framework, while providing the highest-quality LFI maps to date.

Rather than simply probing the degeneracy between instrument calibration and component separation, a better solution

is to actually break it. The optimal approach to do so is by jointly analyzing complementary datasets, each of which provide key information regarding the full system. This insight led to the COSMOGLOBE<sup>1</sup> initiative, which is an Open Source and community-wide effort that aims to derive a single joint model of the radio, microwave, and sub-millimeter sky by combining all available state-of-the-art experiments. An obvious first extension of the LFI-oriented BEYONDPLANCK project is to analyze the *WMAP* measurements in the same framework. Indeed, already as part of the BEYONDPLANCK suite of papers, Watts et al. (2022) integrated *WMAP* *Q*-band time-ordered data (TOD) into the Commander3 framework, calibrated off of the BEYONDPLANCK sky model.

In this paper, we present the first end-to-end Bayesian analysis of the full *WMAP* TOD, processed within the Commander framework. As such, this paper also presents the first ever joint analysis of two major CMB experiments (LFI and *WMAP*) at the lowest possible level, and it therefore constitutes a major milestone of the COSMOGLOBE initiative. In the future, many more datasets will be added, gradually providing stronger and stronger constraints on the true astrophysical sky. Each new experiment will then also in turn improve the instrumental calibration of all previous experiments.

The rest of this paper is organized as follows. In Sect. 2, we provide a brief review of the Bayesian end-to-end statistical framework used in this work, before describing the underlying data and computational expenses in Sect. 3. The main results, as defined by "as expressed by/through?" the global posterior distribution, are described in Sects. 4–6, summarizing instrumental parameters, frequency sky maps, and astrophysical results, respectively. In Sect. 7 we address unresolved issues that should be addressed in future work. We conclude in Sect. 8, and lay a path forward for the COSMOGLOBE project.

## 2. End-to-end Bayesian CMB analysis

The general computational analysis framework used in this work has been described in detail by BeyondPlanck (2022) and Watts et al. (2022) and references therein. In this section, we give a brief summary of the main points, and emphasize in particular the differences with respect to earlier work.

### 2.1. LFI and BEYONDPLANCK

The BEYONDPLANCK project (BeyondPlanck 2022) was the first true application of end-to-end Gibbs sampling in the framework of CMB data analysis. The *Planck* LFI data had been gradually improved through each *Planck* data release, specifically PR1 (Planck Collaboration II 2016), PR2 (Planck Collaboration II 2016), PR3 (Planck Collaboration II 2020), and PR4 (Planck Collaboration II 2016). Even after PR4, however, the final LFI maps still failed significant polarization null tests (in particular for the 44 GHz frequency channel), in particular due to contained poorly measured relative gain uncertainty modes. This sentence is weird. Indeed, already the PR3 analysis it sounds a bit strange to me to have the analysis be the thing that notes something noted a strong degeneracy between the LFI gain solution and the astrophysical foreground model, and to break these the *Planck* LFI Data Processing Center (DPC) implemented an iterative scheme in which the various low-level processing and component separation operations were performed sequentially and repeatedly. However, because these iterations

were performed manually with significant amounts of human interactions, this scheme was very slow, and only four full cycles were completed before the *Planck* collaboration ended. A main motivation for the BEYONDPLANCK project was to automate this process, and perform thousands of such iterations, rather than just four.

This work was highly successful, and the BEYONDPLANCK collaboration was able to generate the first LFI maps with true joint estimation of sky components and underlying instrumental parameters through a statistical process called Gibbs sampling (BeyondPlanck 2022; Galloway et al. 2022a; Basyrov et al. 2022). The code used for this process was called Commander3 (Galloway et al. 2022a). The resulting Gibbs chains allowed for data-driven estimates of the instrumental properties to be determined by exploring the degeneracies between the low-level instrumental parameters themselves and our knowledge of the sky (Andersen et al. 2022; Svalheim et al. 2022b), and in doing so resulted in new state-of-the-art and low-systematics LFI maps.

### 2.2. Official WMAP instrument model and analysis pipeline

The main goal of the current paper is to perform a similar analysis with the *WMAP* time-ordered data, and thereby solve some of the long-standing unresolved issues with the official maps, in particular related to poorly constrained large-scale polarization modes. Before presenting our algorithm, however, it is useful for reference purposes to briefly review the official *WMAP* instrument model and analysis pipeline, which improved gradually over a total of five data releases, often referred to as the 1-, 3-, 5-, 7, and 9-year data releases, respectively. Unless otherwise noted, we will refer to the final 9-year results in the following. The official *WMAP* results delivered by Bennett et al. (2013) will be referred to as *WMAP9*.

The *WMAP* satellite carried forty differential polarization-sensitive radiometers grouped as differencing assemblies (DAs), where one pair was sensitive to the difference in signal at one polarization orientation and the other pair sensitive to the orthogonal polarization. In total, there were ten DAs, which were distributed as one *K*-band (23 GHz), one *Ka*-band (33 GHz), two *Q*-band (41 GHz), two *V*-band (61 GHz), and four *W*-band (94 GHz) DAs. Each radiometer records a science sample every  $1.536/N_{\text{obs}}$  seconds, where  $N_{\text{obs}}$  is 12, 12, 15, 20 and 30 for *K*, *Ka*, *Q*, *V*, and *W*, respectively.

The *WMAP* bandpasses were measured pre-launch on ground, sweeping a signal source through 201 frequencies and recording the output (Jarosik et al. 2003b). The bandpass responses have not been updated since the initial data release. However, as noted by Bennett et al. (2013), there has been an observed drift in the center frequency of *K*, *Ka*, *Q*, and *V*-band corresponding to a  $\sim 0.1\%$  decrease over time. In practice, this did not affect the *WMAP* data processing because each year was mapped separately and co-added afterwards. An effective frequency calculator was delivered in the DR5 release as part of the IDL library to mitigate this effect during astrophysical analyses.<sup>2</sup>

The beams were characterized in the form of maps, with separate products for the near-field and far-field. The main beam and near sidelobes were characterized using a combination of physical optics codes and observations of Jupiter for each horn separately. The maps of Jupiter were then combined with the best-fit physical optics codes to create a map of the beam response (Hill et al. 2009; Weiland et al. 2011; Bennett et al. 2013).

<sup>2</sup> [https://lambda.gsfc.nasa.gov/product/wmap/dr5/m\\_sw.html](https://lambda.gsfc.nasa.gov/product/wmap/dr5/m_sw.html)

<sup>1</sup> <https://cosmoglobe.uio.no>

Far sidelobes were estimated using a combination of laboratory measurements and Moon data taken during the mission (Barnes et al. 2003), as well as a physical optics model described by Hinshaw et al. (2009). To remove the far sidelobe in the TOD, an estimate was calculated by convolving the intensity map and the orbital dipole signal with the measured sidelobe signal (Jarosik et al. 2007). Although the sidelobe pickup was modeled by Barnes et al. (2003), it was determined that the results were small enough to be neglected and have not been explicitly reported in any of the subsequent *WMAP* data releases.

The *WMAP* pointing solution was determined using the bore-sight vectors of individual feedhorns in spacecraft coordinates, in combination with on-board star trackers. Thermal flexure of the tracking structure introduced small pointing errors, as discussed by Jarosik et al. (2007). Using the temperature variation measured by onboard thermistors, the pointing solution was corrected using a model that returns angular deviation per kelvin **What?**. The residual pointing errors were computed using observations of Jupiter and Saturn, and the reported upper limit was given as  $10''$  (Bennett et al. 2013; Greason et al. 2012).

The *WMAP* data were calibrated by jointly estimating the time-dependent gains,  $\mathbf{g}$ , and baselines,  $\mathbf{b}$ , as described by Hinshaw et al. (2007), Hinshaw et al. (2009), and Jarosik et al. (2011). The TOD were initially modeled as having constant gain and baseline for a 1–24 hour period, with parameters that were fit to the orbital dipole assuming  $T_0$  from Mather et al. (1999) and a map made from a previous iteration of the mapmaking procedure. Once the gain and baseline solution had converged, the data were fit to a parametric form of the radiometer response as a function of housekeeping data, given in Appendix A of Greason et al. (2012).

One unique feature of the *WMAP* instrument is that it includes differential pointing. That is, *WMAP* had two primary mirrors positioned on opposite sides of the vertical satellite axis, tilted approximately  $19.5^\circ$  downwards. In total **a strange expression to use here?**, when horn A was pointed at pixel  $p_A$ , horn B was pointed at a pixel  $p_B$  approximately  $141^\circ$  away (Page et al. 2003). The incoming radiation was differenced in the electronics before being deposited on the detectors, essentially recording radiation proportional to  $\mathbf{m}_A - \mathbf{m}_B$  and  $\mathbf{m}_B - \mathbf{m}_A$  (Jarosik et al. 2003b). Each pair of radiometers had a partner pair of radiometers that observed the same pixels with sensitivity to the orthogonal polarization direction. Taking these effects all into account, the total data model for a single radiometer is given by

$$d_t^{\text{imbal}} \propto (1 + x_{\text{im}})T_{p_A} - (1 - x_{\text{im}})T_{p_B} \quad (1)$$

$$= (T_{p_A} - T_{p_B}) + x_{\text{im}}(T_{p_A} + T_{p_B}), \quad (2)$$

where  $T_{p_A}$  and  $T_{p_B}$  are the A- and B-side antenna temperatures. This effect is taken into account during mapmaking. However, inaccuracies in the determination of  $x_{\text{im}}$  will yield a spurious polarization component, and create artificial imbalance modes due to coupling with the sky signal, in particular with the bright Solar CMB dipole (Jarosik et al. 2007). The *WMAP* transmission imbalance factors were fit to the Solar dipole in TOD space, accounting for both common and differential modes (Jarosik et al. 2003a, 2007).

Data were flagged and masked before the final mapmaking step. In particular, station-keeping maneuvers, solar flares, and unscheduled events caused certain data to be unusable – the full catalog of these events is listed in Table 1.8 of Greason et al. (2012). In addition, data were masked depending on the channel frequency and the planet itself, with the full list of exclusion radii enumerated in Table 4 of Bennett et al. (2013).

To create the sky maps  $\mathbf{m}$ , the calibrated data were put into the asymmetric mapmaking equation,

$$\mathbf{P}_{\text{am}}^T \mathbf{N}^{-1} \mathbf{P} \mathbf{m} = \mathbf{P}_{\text{am}}^T \mathbf{N}^{-1} \mathbf{d}, \quad (3)$$

where  $\mathbf{N}$  is the noise covariance matrix, and the pointing matrix  $\mathbf{P}$  is implicitly defined for detectors (**does it make more sense to just say detectors 1 and 2 here? d1 and d2 are strictly speaking referring to datastreams, not the detectors themselves, right?**)  $\mathbf{d}_1$  and  $\mathbf{d}_2$  sensitive to different polarization orientations. For each side A/B, the maps are defined as a function of the Stokes parameters  $T_{A/B}$ ,  $Q_{A/B}$ , and  $U_{A/B}$ , with polarization angle  $\gamma_{A/B}$ , such that

$$\begin{aligned} \mathbf{d}_1 &= \mathbf{P}_1 \mathbf{m} \\ &= (1 + x_{\text{im}})[T_A + Q_A \cos 2\gamma_A + U_A \sin 2\gamma_A + S_A] \\ &\quad + (1 - x_{\text{im}})[-T_B - Q_B \cos 2\gamma_B - U_B \sin 2\gamma_B - S_B], \end{aligned} \quad (4)$$

and

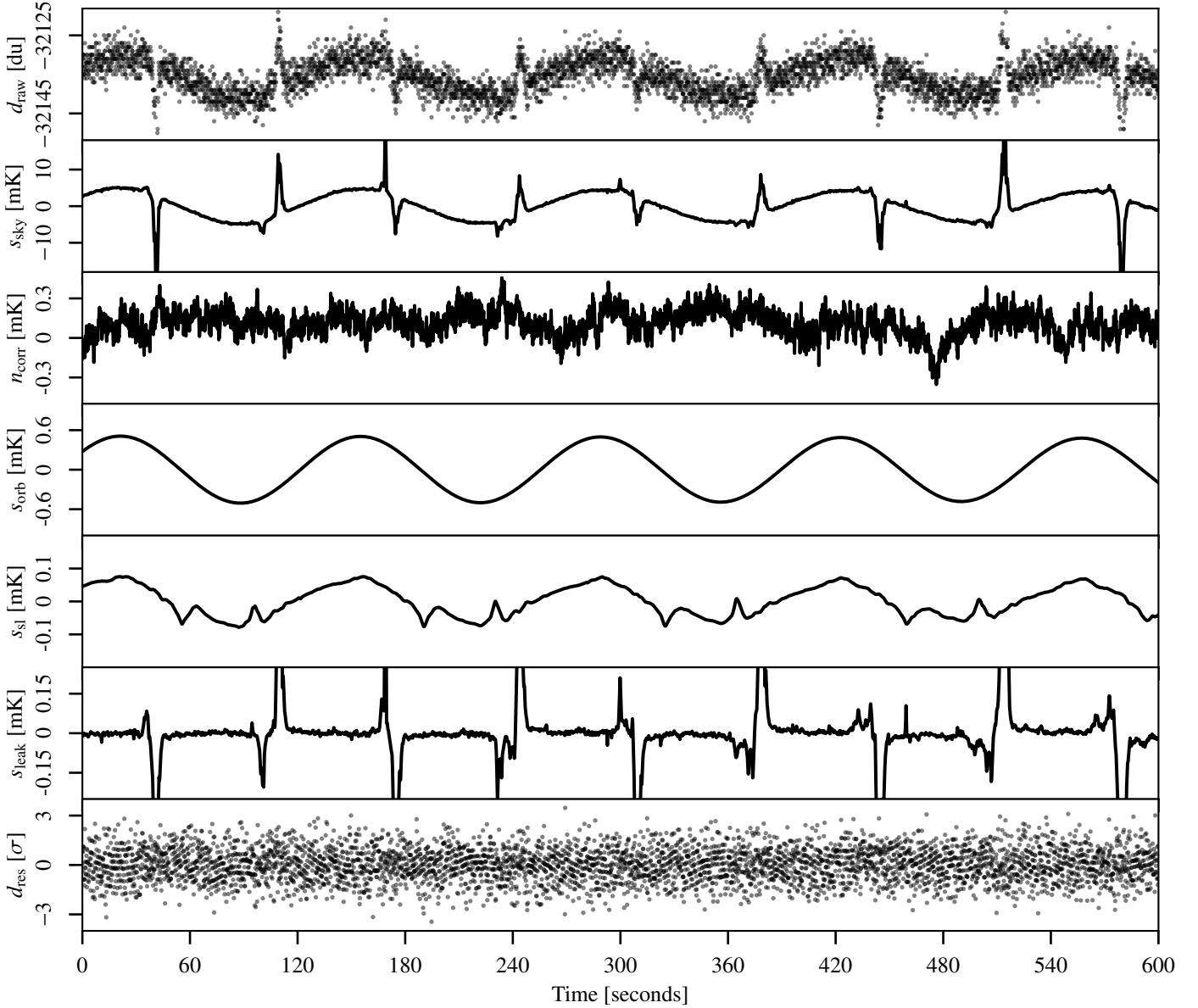
$$\begin{aligned} \mathbf{d}_2 &= \mathbf{P}_2 \mathbf{m} \\ &= (1 + x_{\text{im}})[T_A - Q_A \cos 2\gamma_A - U_A \sin 2\gamma_A - S_A] \\ &\quad + (1 - x_{\text{im}})[-T_B + Q_B \cos 2\gamma_B + U_B \sin 2\gamma_B + S_B]. \end{aligned} \quad (5)$$

In this formalism,  $S$  acts as an extra Stokes parameter that absorbs the effects of differing bandpass responses between radiometers (**same comment here**)  $\mathbf{d}_1$  and  $\mathbf{d}_2$  (Jarosik et al. 2007). The asymmetric mapmaking matrix,  $\mathbf{P}_{\text{am}}$ , was used because, as noted by Jarosik et al. (2011), large signals observed in one beam could leak into the solution for the pixel observed by the other beam, leading to incorrect signals in the final map. The asymmetric mapmaking solution essentially works by only updating the matrix multiplication for beam A when beam A is in a high emission region and beam B is not, and vice versa. Bennett et al. (2013) also identified that these effects are pronounced when there is a steep gradient in the emission across the  $N_{\text{side}} = 512$  pixels within an  $N_{\text{side}} = 16$  superpixel, leading to excesses  $140^\circ$  away from the Galactic center.

An accurate noise model was necessary to perform the maximum likelihood mapmaking, as it required the evaluation of the dense time-space inverse noise covariance matrix  $\mathbf{N}^{-1}$ . The time-space autocorrelation function was estimated for each year of data, which is then Fourier transformed, inverted, and Fourier transformed again to create an effective inverse noise operator  $\mathbf{N}_{tt'}^{-1}$ . Finally, to create the sky maps themselves, the data are treated one year at a time, and the iterative Bi-conjugate Gradient Stabilized Method (BiCG-STAB van der Vorst 1992; Barrett et al. 1994) is applied to the maps.

### 2.3. COSMOGLOBE instrument model

Perhaps the most fundamental difference between the Commander and *WMAP* (and those of most other CMB experiments) analysis pipelines is that while the *WMAP* pipeline models each channel in isolation, the COSMOGLOBE framework simultaneously considers all data, both internally within *WMAP*, and also from all other sources, most notably also including *Planck*. The main advantage of such a global approach is significantly reduced parameter degeneracies, as data from observations with different frequency coverages and instrumental designs break the same degeneracies. For this approach to be computationally tractable, one must establish a global parametric model that simultaneously accounts for



**Fig. 1.** Time-ordered data segment for the K113 radiometer. From top to bottom, the panels show 1) raw uncalibrated TOD  $\mathbf{d}$ ; 2) sky signal  $s_{\text{sky}}$ ; 3) calibrated correlated noise  $n_{\text{corr}}$ ; 4) orbital CMB dipole signal  $s_{\text{orb}}$ ; 5) sidelobe correction  $s_{\text{sl}}$ ; 6) bandpass leakage correction  $s_{\text{leak}}$ ; and 7) residual TOD,  $d_{\text{res}} = (\mathbf{d} - \mathbf{n}_{\text{corr}} - \mathbf{b})/g - s_{\text{sky}} - s_{\text{orb}} - s_{\text{leak}} - s_{\text{sl}}$ , in units of  $\sigma_0[\text{du}]$  for this TOD segment. Note that the vertical range varies significantly from panel to panel.

both the astrophysical sky and all relevant instruments. For the current *WMAP+LFI* oriented analysis, we adopt the following expression (BeyondPlanck 2022),

$$\mathbf{d} = \mathbf{GP}[\mathbf{B}^{\text{symm}} \mathbf{Ma} + \mathbf{B}^{4\pi}(s^{\text{orb}} + s^{\text{fsl}})] + s^{\text{inst}} + \mathbf{n}^{\text{corr}} + \mathbf{n}^{\text{w}}, \quad (7)$$

where  $\mathbf{G}$  is the time-dependent gain in the form of the matrix  $\text{diag}(g_t)$ ;  $\mathbf{P}$  is the  $n_p \times n_t$  pointing matrix, where  $n_p$  is the number of pixels and  $n_t$  the number of TOD datapoints;  $\mathbf{B}^{\text{symm}}$  and  $\mathbf{B}^{4\pi}$  are the symmetrized and full asymmetric beam, respectively;  $\mathbf{M}$  is the mixing matrix between a given component  $c$  with spectral energy distribution  $f_c(v/v_{0,c})$  and a detector  $j$  with bandpass  $\tau_j(v)$ , given by

$$\mathbf{M}_{cj} = \int dv \tau_j(v) f_c(v/v_{c,0}). \quad (8)$$

The maps  $\mathbf{a}$  represent the Stokes parameters for each astrophysical component, while  $s^{\text{orb}}$  is the orbital dipole induced by the

motion of the telescope with respect to the Sun, and  $s^{\text{fsl}}$  is the time-dependent far sidelobe signal. Following Ihle et al. (2022), we model the correlated noise component  $\mathbf{n}^{\text{corr}}$  in terms of a  $1/f$  power spectral density (PSD), which explicitly takes the form  $P_n(f) = \sigma_0^2(1 + (f/f_k)^{\alpha})$ , where  $\sigma_0$  denotes the white noise amplitude,  $f_{\text{knee}}$  is the so-called  $1/f$  knee frequency, and  $\alpha$  is a free power law slope. For notational purposes, we denote the set of all correlated noise parameters by  $\xi_n = \{\sigma_0, f_k, \alpha\}$ . We note that this model represents a significant approximation, as the actual *WMAP* noise is known to exhibit a significantly colored noise at high temporal frequencies. The main impact of this approximation is a worse-than-expected  $\chi^2$  goodness of fit statistic. However, measured in absolute noise levels the effect is very small, and has very little if any impact on the final science results; for further discussion of this approximation, see Sect. 7.1.

The term  $s^{\text{inst}}$  denotes any instrument-specific terms that might be required for a given experiment. For instance, for LFI

it is used to model the 1 Hz spike contribution due to electronic cross-talk. For *WMAP*, we use it for first-order baseline corrections, and set  $s_{\text{corr}}^{\text{WMAP}} = b_0 + b_1 t$ , where  $b_0$  and  $b_1$  represent the mean and slope of the baselines over the data segment in question. We note that while the *WMAP* team fitted a single constant baseline over either 1- or 24-hour periods, our data segments are typically about 3 days long (corresponding to a number of samples that is an exact power of 2 to optimize Fourier transforms). A natural question is therefore whether non-linear baseline variations could induce artefacts. In this regard, it is important to note that the correlated noise component effectively acts as a single-sample baseline correction that can absorb by far most such non-linearities, as long as their total effect on the power spectrum does not exceed that imposed by the  $1/f$  model. In practice, that is a very mild constraint. At the same time, visual inspection of  $\mathbf{n}^{\text{corr}}$  projected into sky maps provides a very powerful check on any potential baseline residuals; these will appear as correlated stripes aligned with the *WMAP* scanning path. In sum, it is important to note that COSMOGLOBE model allows for a more flexible baseline behaviour than the *WMAP* pipeline.

A third notable difference between the *WMAP* and COSMOGLOBE data models concerns bandpass mismatch. While the *WMAP* pipeline simply projects out any bandpass difference from the polarization maps by solving for the spurious  $S$  maps, we model it explicitly through the use of the global astrophysical sky model (Svalheim et al. 2022a). Explicitly, the expected calibrated sky signal for radiometer  $j$  is given by

$$m_{p,j} = \mathbf{B}_{p,p'} \sum_c \mathbf{M}_{c,j} a_p^c + n_{j,p}^w. \quad (9)$$

Since  $\mathbf{M}_{c,j}$  encodes the bandpass response of every detector  $j$  to every sky component  $c$ , the detector-specific maps,  $\mathbf{m}_j$ , will each be slightly different depending on their bandpass  $\tau_j$ . Therefore, before averaging different detectors together, we estimate the average over all detectors in a given frequency channel  $\mathbf{m} \equiv \langle \mathbf{m}_j \rangle$ , and subtracts it directly in the timestream;

$$\delta s_{t,j}^{\text{leak}} = \mathbf{P}_{t,p}^j \mathbf{B}_{p,p'}^j (\mathbf{m}_{j,p'} - \mathbf{m}_{p'}). \quad (10)$$

This leakage term uses the expected bandpass response to remove the expected component that deviates from the mean in the timestream, directly reducing polarization contamination.

To build intuition regarding this model, we plot in Fig. 1 both the TOD and the individual model components for an arbitrarily selected ten-minute segment for the *WMAP*'s K113 radiometer. The uncalibrated data,  $\mathbf{d}_{\text{raw}}$ , is displayed in the top panel, with the sky signal  $\mathbf{s}_{\text{sky}} = \mathbf{P}\mathbf{B}^{\text{symm}}\mathbf{M}\mathbf{a}$  plotted directly underneath. The next four panels show the correlated noise realization  $\mathbf{n}_{\text{corr}}$ , the orbital dipole  $\mathbf{s}_{\text{orb}}$ , the far sidelobe contribution  $\mathbf{s}_{\text{sl}}$ , and the bandpass leakage  $\mathbf{s}_{\text{leak}}$ . Finally, we also plot the time-ordered residual for this segment of data, obtained by subtracting the model from the raw data, in units of the estimated white noise level.

#### 2.4. Sky model

Following BeyondPlanck (2022), we assume that the sky (as defined by  $\mathbf{M}\mathbf{a}$  in Eq. 7) across the *WMAP* frequencies can be modeled as a linear combination of CMB fluctuations ( $\mathbf{a}_{\text{CMB}}$ ), synchrotron ( $\mathbf{a}_s$ ), free-free emission ( $\mathbf{a}_{\text{ff}}$ ), anomalous microwave emission ( $\mathbf{a}_{\text{ame}}$ ), thermal dust ( $\mathbf{a}_d$ ), and radio point sources ( $\mathbf{a}_{j,\text{src}}$ ). Explicitly, we assume that the astrophysical sky (in units

of brightness temperature) may be modelled as follows,

$$s_{\text{RJ}} = (\mathbf{a}_{\text{CMB}} + \mathbf{a}_{\text{quad}}(\nu)) \frac{x^2 e^x}{(e^x - 1)^2} + \quad (11)$$

$$+ \mathbf{a}_s \left( \frac{\nu}{\nu_{0,s}} \right)^{\beta_s} + \quad (12)$$

$$+ \mathbf{a}_{\text{ff}} \left( \frac{\nu_{0,\text{ff}}}{\nu} \right)^2 \frac{g_{\text{ff}}(\nu; T_e)}{g_{\text{ff}}(\nu_{0,\text{ff}}; T_e)} + \quad (13)$$

$$+ \mathbf{a}_{\text{ame}} e^{\beta_{\text{AME}}(\nu - \nu_{0,\text{ame}})} + \quad (14)$$

$$+ \mathbf{a}_d \left( \frac{\nu}{\nu_{0,d}} \right)^{\beta_d + 1} \frac{e^{h\nu_{0,d}/k_B T_d} - 1}{e^{h\nu/k_B T_d} - 1} + \quad (15)$$

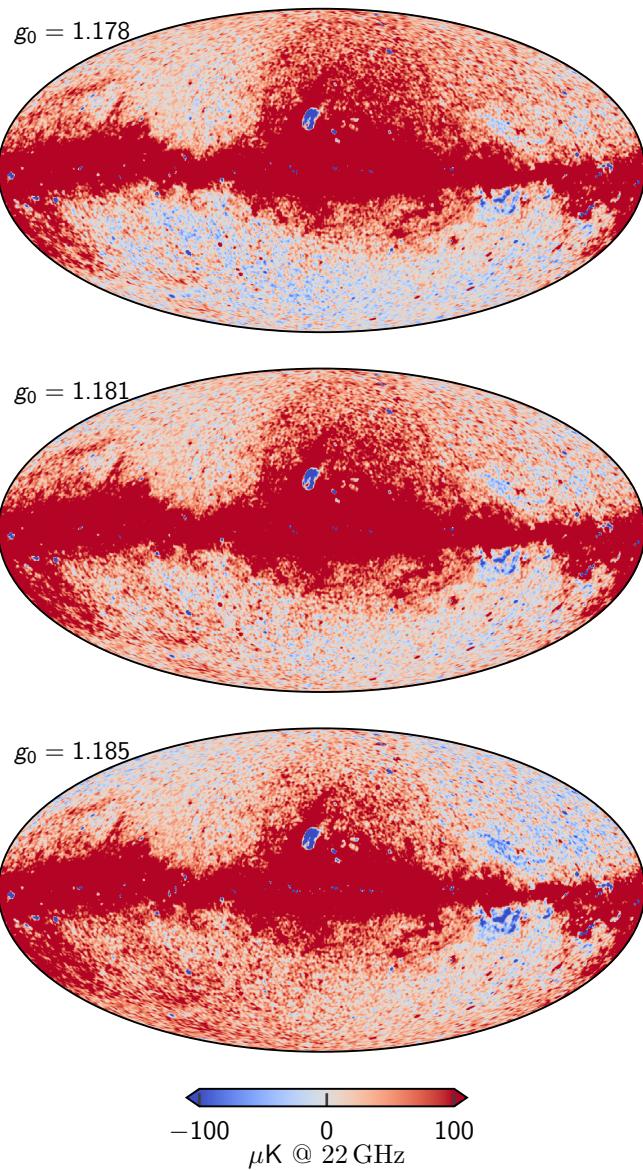
$$+ U_{\text{mJy}} \sum_{j=1}^{N_{\text{src}}} \mathbf{a}_{j,\text{src}} \left( \frac{\nu}{\nu_{0,\text{src}}} \right)^{\alpha_{j,\text{src}} - 2}, \quad (16)$$

where  $x = h\nu/kT_{\text{CMB}}$ ;  $\nu_{0,i}$  is a reference frequency for component  $i$ ;  $\beta_s$  is a power-law index for synchrotron emission (which may take different values for temperature and polarization);  $T_e$  is the electron temperature, and  $g_{\text{ff}}$  is the so-called Gaunt factor (Dickinson et al. 2003);  $\beta_{\text{AME}}$  is an exponential scale factor for AME emission (see below);  $\beta_d$  and  $T_d$  are the emissivity and temperature parameters for a single modified blackbody thermal dust model;  $\alpha_{j,\text{sr}}$  is the spectral index of point source  $j$  relative to the same source catalog as used by Planck Collaboration IV (2018); and  $U_{\text{mJy}}$  is the conversion factor between flux density (in millijansky) and brightness temperature (in  $K_{\text{RJ}}$ ) for the channel in question. Finally,  $\mathbf{a}_{\text{quad}}$  accounts for a relativistic quadrupole correction due to the Earth's motion through space (Notari & Quartin 2015).

In general, this model is nearly identical to the one adopted by BeyondPlanck (2022). However, there is one notable exception, namely the spectral energy density (SED) for the AME component,  $s_0^{\text{sd}}(\nu)$ . In this current work, we adopt a simple exponential function for this component, as for instance proposed by Hensley et al. (2015), which is notably different from the SpDust2 model (Ali-Haïmoud et al. 2009; Ali-Haïmoud 2010; Silsbee et al. 2011) that was used in the BEYONDPLANCK analysis. The motivation for this modification is discussed in detail by Watts et al. (2023): First and foremost, the current combination of *WMAP* and LFI data appears to prefer a higher signal amplitude at frequencies between 40 and 60 GHz than can easily be supported by SpDust2. This was first noted by Planck Collaboration IX (2016), who solved this issue by introducing a second independent AME component. For the original BEYONDPLANCK analysis, on the other hand, this excess was not statistically significant, simply because that analysis did not include the powerful *WMAP* K-band data. In the current analysis, the excess is obvious. The observation that a simple one-parameter exponential model fits the data as well as the complicated multi-parameter model of Planck Collaboration IX (2016) is a novel result from the current work. Indeed, it fits also about as well as the commonly used log-normal model derived by Stevenson (2014), which also has one extra parameter. By virtue of having fewer degrees of freedom than any of the previous models, we adopt the exponential model in the following.

#### 2.5. Priors and poorly measured modes

The model described in Sects. 2.2 and 2.4 is prone to several degeneracies, allowing for unphysical solutions to be explored in the Gibbs chain. Such unphysical degeneracies are highly undesirable for two main reasons. First, they increase the statisti-



**Fig. 2.** Dependence on AME amplitude evaluated at 22 GHz as a function of absolute calibration. Each map comes from the fifth iteration of a dedicated `Commander` run that fixed  $g_0$  while letting all other TOD parameters be fit. The values of  $g_0 = 1.178$  and  $g_0 = 1.185$  represent  $3.5\sigma$  draws from the prior distribution with mean 1.1815 and standard deviation 0.001. The dipole visible in the top and bottom panels is aligned perfectly with the Solar dipole, and is directly due to variations in the K-band absolute calibration.

cal uncertainties on most (if not all) other important parameters in the model – sometimes to the point that the target quantity is rendered entirely unmeasurable. Secondly, and perhaps even more importantly, the data model described above is known to be a (sometimes crude) approximation to the real observations, and there will invariably be modelling errors. Degeneracies then generally tend to amplify their impact, in the sense that any unconstrained parameters will typically be used to fit such small modelling errors. For both these reasons, it is generally preferable to impose either informative or algorithmic priors on the unconstrained parameters, rather than to leave them entirely unconstrained in the model.

An important example of an algorithmic prior is the foreground smoothing prior used by [Planck Collaboration IV \(2018\)](#) and [Paradiso et al. \(2022\)](#), which dictates that astrophysical foregrounds has to be smooth on small angular scales. This is justified by noting that the angular spectrum on large and intermediate scales typically falls as a power-law in multipole space; extrapolating this into the noise dominated regime prevents the overall foreground model from becoming degenerate at small scales.

Correspondingly, important examples of informative priors are the use of HFI constraints on the thermal dust SED parameters,  $\beta_d$  and  $T_d$  in BEYONDPLANCK. Because that analysis only included the highest HFI frequency channels, they had very little constraining power on the thermal dust SED. Rather than trying to fit these directly from LFI WMAP alone, they instead imposed informative Gaussian priors on each of these parameters, as derived from the HFI observations ([Planck Collaboration IV 2018](#)).

In the current analysis, and unless otherwise noted, we adopt the same algorithmic and informative priors as [BeyondPlanck \(2022\)](#). However, there are three notable exceptions, as detailed below, all of which are dictated either by the fact that we include the WMAP *K*-band channel (which has a strong impact on the low-frequency foreground model), or by the fact that we now process the WMAP data in time-domain, and therefore are subject to the same degeneracies as the official WMAP low-level pipeline, and that were solved with similar implicit or explicit prior in the original analysis.

First and foremost, and as detailed in Sect. 7.3, we observe in the current analysis a very strong degeneracy between the absolute calibration of the *K*-band channel and the dipole of the AME map. This makes intuitive sense, since *K*-band is by far the strongest channel in terms of AME signal-to-noise ratio, exceeding that of LFI 30 GHz by about a factor of four; see Sect. 6.3. Effectively, a small variation in the absolute gain may be countered by subtracting corresponding CMB Solar dipole variation from the AME map, and end up with a nearly identical total  $\chi^2$ ; the orbital CMB dipole is not bright enough at 23 GHz relative to AME emission to break this degeneracy.

This is illustrated in Fig. 2, which shows the derived AME amplitude map for three different values of the mean *K*-band gain,  $g_0$ , namely 1.178, 1.181, and 1.184  $\mu\text{mK}^{-1}$ ; the extreme values differ only by 0.25 %. All of these three values appear equally acceptable from a pure  $\chi^2$  point-of-view, relative to the noise level and modeling errors of these data. At the same time, it is clear from visual inspection that only the middle value actually makes physical sense, as compared to what we know about the structure of the Milky Way. For this reason, we apply a Gaussian prior on the absolute *K*-band gain of  $g_0 \sim \mathcal{N}(1.181, 0.001^2)$  in the following, to regularize this issue. Thus, the extreme panels in Fig. 2 represent  $\pm 2.5\sigma$  outliers, respectively, and will appear in our Markov chains with a frequency of about 1-in-100.

It is reasonable to ask why the WMAP pipeline produced sensible results without applying such a prior during their calibration procedure. We posit that the answer is due to the main difference between the two approaches. While COSMOGLOBE attempts to fit a single overall parametric model to all data at once, the WMAP pipeline calibrated each channel independently by co-adding data from one channel into a map, subtracting that map from the TOD, fitting the gain to the orbital dipole, and iterating until the solution became stable. An advantage of the single-channel approach is that the solution is independent of the assumed sky model. However, a disadvantage is that it is impossible to break any potential inherent degeneracies; it cannot be combined with external observations in any meaningful way.

One important example of this for the *WMAP* data is a strong degeneracy between the transmission imbalance factors and the polarized sky signal; it is exceedingly difficult to break this degeneracy using data from only one radiometer alone, and the resulting errors will propagate to most other aspects in the analysis. In the global approach, on the other hand, the polarization modes that are poorly measured by *WMAP* alone are well measured by *Planck*, and the result is an overall better fit.

Second, as reported by Svalheim et al. (2022b) for the BEYONDPLANCK analysis, another important degeneracy in the current global model concerns the spectral index of polarized synchrotron emission versus the time-variable detector gain; when fitting both the polarized synchrotron amplitude and calibration freely without priors, the synchrotron spectral index at high Galactic latitudes tend to be biased toward unreasonably flat values,  $\beta \lesssim -2.5$ , which was likely due to a low level of unmodelled systematics, for instance temperature-to-polarization leakage, rather than true polarized synchrotron emission. In turn, this resulted in a clearly contaminated CMB sky map with a strong synchrotron morphology. To break this degeneracy, Svalheim et al. (2022b) chose to marginalize the high-latitude synchrotron spectral index over a Gaussian prior of  $\mathcal{N}(-3.30, 0.1^2)$ , informed by Planck Collaboration V (2020), rather than estimate it from the data themselves. In the current analysis, we observe the same degeneracy, and the introduction of the  $K$ -band data is not sufficient to break it on its own. For this reason, we choose to apply the same informative prior in the current analysis.

Third and finally, we also marginalize over the AME scale index with a prior of  $\beta_{\text{AME}} \sim \mathcal{N}(3.56, 0.1^2)$ . The parameters of these priors were determined by running a grid over  $\beta_{\text{AME}}$ , and identifying the range that resulted in reasonable residuals near the Galactic plane, similar to that shown in Fig. 2 for the absolute calibration of  $K$ -band. We note that this prior should in principle be replaced with direct  $\chi^2$ -based posterior optimization, combined with a properly tailored analysis mask. However, the recent release of the QUIJOTE data (Rubiño-Martín et al. 2023), which covers the 11–19 GHz frequency range, suggests that the entire AME model should be revisited in a future joint *WMAP*+LFI+QUIJOTE analysis. We therefore leave detailed prior and SED optimization to that work. For further information regarding AME modelling with the current dataset, we refer the interested reader to Watts et al. (2023).

## 2.6. Posterior distribution and Gibbs sampling

As shown by BeyondPlanck (2022), this joint parametric description of the instrumental effects and sky allows us to write down a total model for the data,  $\mathbf{d} = \mathbf{s}^{\text{tot}}(\omega) + \mathbf{n}^{\text{w}}$ , where  $\mathbf{s}^{\text{tot}}$  encompasses all of the terms in Eq. (7) except for the white noise term. Assuming that all instrumental effects have been modeled adequately, and that the white noise is Gaussian distributed, the data should then also be Gaussian distributed with a mean of  $\mathbf{s}^{\text{tot}}(\omega)$  and variance  $\sigma_0^2$ . In general, the likelihood reads

$$P(\mathbf{d} | \omega) \propto \exp\left(-\frac{1}{2} \sum_t \frac{(d_t - s_t^{\text{tot}}(\omega))^2}{\sigma_0^2}\right). \quad (17)$$

If  $\mathbf{d} \sim \mathcal{N}(\mathbf{s}^{\text{tot}}, \sigma_0^2)$  is the correct model for the data, the argument of the exponent is proportional to a  $\chi^2$ -distribution with  $n_{\text{TOD}}$  degrees of freedom. In the limit of large  $n$ , a  $\chi^2$  distribution is well-approximated by a Gaussian with mean  $n$  and variance  $2n$ . Therefore we define and use in the following the reduced nor-

malized  $\chi^2$  statistic,

$$\chi^2 \equiv \frac{\sum_t ((d_t - s_t^{\text{tot}})^2 / \sigma_0^2) - n_{\text{TOD}}}{\sqrt{2n_{\text{TOD}}}}, \quad (18)$$

which is approximately drawn from the standard normal distribution  $\mathcal{N}(0, 1)$ .

Following BeyondPlanck (2022), the COSMOGLOBE Gibbs chain for this analysis is given by

$$\mathbf{g} \leftarrow P(\mathbf{g} \mid \mathbf{d}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, C_\ell, \theta) \quad (19)$$

$$\mathbf{n}^{\text{corr}} \leftarrow P(\mathbf{n}^{\text{corr}} \mid \mathbf{d}, \mathbf{g}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, C_\ell, \theta) \quad (20)$$

$$\xi_n \leftarrow P(\xi_n \mid \mathbf{d}, \mathbf{g}, \mathbf{n}^{\text{corr}}, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, C_\ell, \theta) \quad (21)$$

$$\mathbf{s}^{\text{inst}} \leftarrow P(\mathbf{s}^{\text{inst}} \mid \mathbf{d}, \mathbf{g}, \mathbf{n}^{\text{corr}}, \xi_n, \boldsymbol{\beta}, \mathbf{a}, C_\ell, \theta) \quad (22)$$

$$\boldsymbol{\beta} \leftarrow P(\boldsymbol{\beta} \mid \mathbf{d}, \mathbf{g}, \mathbf{n}^{\text{corr}}, \xi_n, \mathbf{s}^{\text{inst}}, C_\ell, \theta) \quad (23)$$

$$\mathbf{a} \leftarrow P(\mathbf{a} \mid \mathbf{d}, \mathbf{g}, \mathbf{n}^{\text{corr}}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, C_\ell, \theta) \quad (24)$$

$$C_\ell \leftarrow P(C_\ell \mid \mathbf{d}, \mathbf{g}, \mathbf{n}^{\text{corr}}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, \theta) \quad (25)$$

$$\theta \leftarrow P(\theta \mid \mathbf{d}, \mathbf{g}, \mathbf{n}^{\text{corr}}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, C_\ell), \quad (26)$$

with each step requiring its own dedicated sampling algorithm. The Commander3 pipeline is designed so that results of each Gibbs sample can be easily passed to each other, and that the internal calculations of each step do not directly depend on the inner workings of each other, which greatly increases modularity of the code.

## 2.7. Sampling algorithms

Before we discuss the results of this Gibbs chain as applied to the *Planck* LFI and *WMAP* data, we summarize the TOD processing steps in this section. Each step of the Gibbs chain requires its own conditional distribution sampling algorithm. In Sect. 2.7.1 we review the sampling algorithms implemented in the BEYONDPLANCK suite of papers, while Sects. 2.7.2–2.7.4 provide an overview of the *WMAP*-specific processing steps.

### 2.7.1. Review of sampling algorithms

By far most of the techniques required for *WMAP* data analysis have already been described in the BEYONDPLANCK project and implemented in Commander3. This section includes a summary of the algorithms that were used previously for the analysis of LFI data. In each of these cases, every part of the model not explicitly mentioned is held fixed unless specified otherwise.

Noise estimation and calibration are described by Ihle et al. (2022) and Gjerløw et al. (2022), respectively. As noted in those works, these two steps are strongly correlated, simply because the timestream

$$d_{t,i} = g_{q,i} s_{t,i}^{\text{tot}} + n_{t,i}^{\text{corr}} + n_{t,i}^{\text{wn}} \quad (27)$$

may be almost equally well fit by two solutions defined by  $g' = g s^{\text{tot}} / (s^{\text{tot}})'$  or  $(n^{\text{corr}})' = n^{\text{corr}} + g s^{\text{tot}} + g'(s^{\text{tot}})'$ ; the only thing that breaks this degeneracy is the noise PSD, which is a relatively loose constraint. A Gibbs sampler is by nature not very effective for nearly degenerate distributions, and we therefore instead define a joint sampling step for the correlated noise and gain. In practice, this is done by first drawing the calibration from its marginal distribution with respect to  $n^{\text{corr}}$ , and then drawing  $n^{\text{corr}}$  from its conditional distribution with respect to  $g$ ,

$$\mathbf{g} \leftarrow P(\mathbf{g} \mid \mathbf{d}, \xi_n, \dots) \quad (28)$$

$$\mathbf{n}^{\text{corr}} \leftarrow P(\mathbf{n}^{\text{corr}} \mid \mathbf{d}, \mathbf{g}, \xi_n, \dots). \quad (29)$$

It is easy to see that this is a valid sample from the joint distribution simply from the definition of a conditional distribution,  $P(\mathbf{g}, \mathbf{n}^{\text{corr}} | \omega) = P(\mathbf{n}^{\text{corr}} | \mathbf{g}, \omega)P(\mathbf{g} | \omega)$ . In practice, this simply means that when sampling for  $\mathbf{g}$ , the covariance matrix  $\mathbf{N} = \mathbf{N}_{\text{wn}} + \mathbf{N}_{\text{corr}}$  must be used, rather than just  $\mathbf{N}_{\text{wn}}$ .

**Commander3** models the gain at each timestream  $t$  for a detector  $i$  as

$$\mathbf{g}_{t,i} = g_0 + \Delta g_i + \delta g_{q,i} \quad (30)$$

where  $q$  labels the time interval for which we assume the gain is constant, typically a single scan. In order to sample the gain, we write down a generative model for the TOD,

$$\mathbf{d}_i = \mathbf{g}_i s_i^{\text{tot}} + \mathbf{n}_i^{\text{tot}} \sim \mathcal{N}(\mathbf{g}_i s_i^{\text{tot}}, \mathbf{N}_i). \quad (31)$$

Since the  $\mathbf{d}_i$  is given as a linear combination of the fixed signal and the gains, a random sample of the gain can be drawn by solving<sup>3</sup>

$$[(\mathbf{s}_i^{\text{tot}})^T \mathbf{N}_i^{-1} \mathbf{s}_i^{\text{tot}}] \mathbf{g}_i = (\mathbf{s}_i^{\text{tot}})^T \mathbf{N}_i^{-1} \mathbf{d}_i + (\mathbf{s}_i^{\text{tot}})^T \mathbf{N}_i^{-1/2} \boldsymbol{\eta}. \quad (32)$$

Note that the  $\mathbf{N}_i$  depends implicitly on the noise PSD  $\xi_n$ , while the specific realization of  $\mathbf{n}^{\text{corr}}$  is accounted for in the covariance matrix. As detailed by Gjerløw et al. (2022), Commander3 samples in practice  $g_0$ ,  $\Delta g_i$ , and  $\delta g_{q,i}$  in separate sampling steps. Specifically, the absolute calibration  $g_0$  is for the CMB-dominated channels only fitted using the orbital dipole, while the relative calibrations,  $\Delta g_i$ , exploits the full sky signal. The same is true for the time-dependent gain fluctuations,  $\delta g_{q,i}$ , and in this case an additional smoothness prior is applied through an effective Wiener filter. It is worth noting that the Gibbs chain is formally broken by fitting the absolute gain  $g_0$  to the orbital dipole alone, as opposed to the full sky signal. However, this makes the sampling more robust with respect to unmodeled systematic effects, somewhat analogous to applying a confidence mask when estimating the CMB power spectrum.

The correlated noise sampling, described by Ihle et al. (2022), follows a similar procedure, except this now conditions upon the previous gain estimate, which is sampled immediately before the correlated noise component in the code. Similar to the gain case, we can write a generative model for the data,

$$\mathbf{d}_i = \mathbf{g}_i s_i^{\text{tot}} + \mathbf{n}_i^{\text{corr}} + \mathbf{n}_i^{\text{wn}} \sim \mathcal{N}(\mathbf{g}_i s_i^{\text{tot}}, \mathbf{N}_{\text{corr},i} + \mathbf{N}_{\text{wn},i}). \quad (33)$$

Given fixed  $\mathbf{r}_i = \mathbf{d}_i - \mathbf{g}_i s_i^{\text{tot}}$ , we can again write a sampling equation,

$$(\mathbf{N}_{\text{corr},i}^{-1} + \mathbf{N}_{\text{wn},i}^{-1}) \mathbf{n}_i^{\text{corr}} = \mathbf{N}_{\text{wn},i}^{-1} \mathbf{r}_i + \mathbf{N}_{\text{wn},i}^{-1/2} \boldsymbol{\eta}_1 + \mathbf{N}_{\text{corr},i}^{-1/2} \boldsymbol{\eta}_2. \quad (34)$$

This gives a sample of the underlying correlated noise.

To sample the correlated noise parameters, we assume that the correlated noise is drawn from a correlated Gaussian and form the conditional posterior distribution,

$$P(\xi_n | \mathbf{n}^{\text{corr}}) \propto \frac{\exp[-\frac{1}{2}(\mathbf{n}^{\text{corr}})^T \mathbf{N}_{\text{corr}}^{-1} \mathbf{n}^{\text{corr}}]}{\sqrt{|\mathbf{N}_{\text{corr}}|}} P(\xi_n). \quad (35)$$

The simplest and most commonly used parameterization for correlated noise is given by

$$\mathbf{N}_{\text{corr}}(f) = \sigma_0^2 \left( \frac{f}{f_{\text{knee}}} \right)^\alpha. \quad (36)$$

<sup>3</sup> See, e.g., Appendix A.2 of BeyondPlanck (2022) for a derivation of this result.

This can in principle be modified, and for *Planck* LFI a Gaussian log-normal bump was added at a late stage in the BEYOND-PLANCK analysis. Rather than sampling for  $\sigma_0$ , we effectively fix the white noise level to the noise level at the highest frequency, e.g.,

$$\sigma_0^2 \equiv \frac{\text{Var}(r_{t+1} - r_t)}{2}, \quad (37)$$

where  $t$  and  $t + 1$  are consecutive time samples, and  $\mathbf{r} \equiv \mathbf{d} - \mathbf{gs}^{\text{tot}} - \mathbf{n}^{\text{corr}}$ . In practice, this makes  $\sigma_0$  a deterministic function of the sampled sky and gain parameters. The parameters  $\alpha$  and  $f_{\text{knee}}$  are not linear in the data, and they can be sampled efficiently using a standard inversion sampler (see, e.g., Appendix A.3 of BeyondPlanck (2022) or Chapter 7.3.2 of Press et al. (2007) for further details). In practice, this requires computing the posterior over a linear grid one parameter at a time.

Once the instrumental parameters have been sampled, Commander3 computes the calibrated TOD for each band,

$$r_{t,j} = \frac{d_{t,j} - n_{t,j}^{\text{corr}}}{g_{t,j}} - (s_{t,j}^{\text{orb}} + s_{t,j}^{\text{fsi}} + \delta s_{t,j}^{\text{leak}} + s_{t,j}^{\text{inst}}) \quad (38)$$

where  $s^{\text{orb}}$  is the orbital dipole (Gjerløw et al. 2022),  $s^{\text{fsi}}$  is the far sidelobe timestream (Galloway et al. 2022b),  $\delta s^{\text{leak}}$  is the bandpass leakage (Svalheim et al. 2022a), and  $s^{\text{inst}}$  is some instrumental-specific contribution, e.g., the 1 Hz electronic spike for LFI. With a correlated noise realization removed, one can perform simple binned mapmaking, weighting each pixel by the white noise amplitude.

## 2.7.2. Differential mapmaking

The first additional algorithm that needs to be added to Commander3 in order to process WMAP TOD data is support for differential mapmaking (Watts et al. 2022). After calibration and correction for instrumental effects, the TOD can be modeled as

$$\mathbf{d} = \mathbf{Pm} + \mathbf{n}^{\text{w}}, \quad (39)$$

where

$$\mathbf{m} = \mathbf{B}^{\text{symm}} \mathbf{Ma} \quad (40)$$

is the expected map for each detector after removing the orbital dipole, far sidelobe, baseline, and a realization of correlated noise. The differential pointing strategy can be represented in matrix form as

$$\begin{aligned} \mathbf{P}_{tp} = & (1 + x_{\text{im}})(\delta_{p' p'_A} + \delta_{p' p'_B} \cos 2\psi_A + \delta_{p' p'_A} \sin 2\psi_A) \\ & - (1 - x_{\text{im}})(\delta_{p' p'_B} - \delta_{p' p'_B} \cos 2\psi_B - \delta_{p' p'_B} \sin 2\psi_B) \end{aligned} \quad (41)$$

where  $p_A$  and  $p_B$  are the time-dependent pointings for each DA. The maximum likelihood map can now in principle be derived using the usual mapmaking equation,

$$\mathbf{P}^T \mathbf{N}^{-1} \mathbf{Pm} = \mathbf{P}^T \mathbf{N}^{-1} \mathbf{d}. \quad (42)$$

For a single-horn experiment, i.e., *Planck* LFI, this reduces to a  $3 \times 3$  matrix that can be inverted for each pixel independently. For the pointing matrix in Eq. (41), this is no longer possible, as there is inherently coupling between horns A and B in the timestreams. The  $3N_{\text{pix}} \times 3N_{\text{pix}}$  matrix can be solved using an iterative algorithm, e.g., preconditioned conjugate gradients (Shewchuk 1994).

Jarosik et al. (2011) identified an issue where a large difference in the sky temperature values at pixel A versus pixel B induced artifacts in the mapmaking procedure. We adopt the procedure first described by Hinshaw et al. (2003) where only the pixel in a bright region, defined by a small processing mask (Bennett et al. 2013) is accumulated, thus modifying the mapmaking equation to

$$\mathbf{P}_{\text{am}}^T \mathbf{N}^{-1} \mathbf{P} \mathbf{m} = \mathbf{P}_{\text{am}}^T \mathbf{N}^{-1} \mathbf{d}. \quad (43)$$

This equation can be solved using the BiCG-STAB algorithm for a non-symmetric matrix  $\mathbf{A}$  where  $\mathbf{Ax} = \mathbf{b}$ . We apply a preconditioner  $\mathbf{M}$  by numerically inverting the same problem with  $N_{\text{side}} = 16$  maps and applying a diagonal noise matrix. Numerically, we define convergence as when the residual  $\mathbf{r} \equiv \mathbf{b} - \mathbf{Ax}$  satisfies  $\mathbf{r}^T \mathbf{M}^{-1} \mathbf{r} / \mathbf{b}^T \mathbf{M}^{-1} \mathbf{b} < 10^{-10}$ , which typically takes about 20 iterations for producing frequency maps.

### 2.7.3. Transmission imbalance estimation

Transmission imbalance, the differential power transmission of the optics and waveguide components, can be parameterized as

$$d_{t,j} = g_{t,j}[(1 + x_{\text{im},j})s_{t,j}^{\text{tot,A}} - (1 - x_{\text{im},j})s_{t,j}^{\text{tot,B}}] + n_t. \quad (44)$$

This can be decomposed into a differential (d) and common-mode (c) signal such that

$$d_{t,j} = g_{t,j}[s_{t,j}^{\text{d}} + x_{\text{im},j}s_{t,j}^{\text{c}}] + n_t. \quad (45)$$

In this form, the imbalance parameters can be estimated by drawing Gaussian samples from the standard mean and standard deviation over the entire mission. To draw samples for  $x_{\text{im},j}$ , we construct a sampling routine analogous to the gain estimation of Eq. (32) and correlated noise estimation of (34), with  $\mathbf{r} = \mathbf{d} - \mathbf{gs}^{\text{d}}$ ,

$$[(\mathbf{gs}^{\text{c}})^T \mathbf{N}^{-1} \mathbf{gs}^{\text{c}}]x_{\text{im}} = (\mathbf{gs}^{\text{c}})^T \mathbf{N}^{-1} \mathbf{r} + (\mathbf{gs}^{\text{c}})^T \mathbf{N}^{-1/2} \boldsymbol{\eta}, \quad (46)$$

essentially cross-correlating the common-mode signal with  $\mathbf{r}$  with appropriate weights and adding a Gaussian random variable with the correct weighting. Note that we are marginalizing over the correlated noise here by using  $\mathbf{N} = \mathbf{N}_{\text{wn}} + \mathbf{N}_{\text{corr}}$ . This mitigates any baseline drifts being erroneously attributed to the common-mode signal and biasing the estimate of  $x_{\text{im}}$ .

The WMAP procedure, described by Jarosik et al. (2003a), fit for common-mode and differential coefficients along with a cubic baseline over 10 precession periods at a time, corresponding to 10 hours of observation. The mean and uncertainty were then calculated by averaging and taking the standard deviation of these values. This approach has the benefit of allowing for the tracking of possible transmission imbalance variation throughout the mission. However, none of the WMAP suite of papers have found evidence for this, and it has not arisen in our analysis, so we model this as an effect whose value is constant throughout the mission.

### 2.7.4. Baseline sampling

The data model adopted by Hinshaw et al. (2003) can be written in raw digital units (du) as

$$\mathbf{d} = \mathbf{GPBM} \mathbf{a} + \mathbf{n} + \mathbf{b}, \quad (47)$$

where  $\mathbf{b}$  is the instrumental baseline and  $\mathbf{n}$  is the total instrumental noise. As noted above, Commander3 divides the noise into

**Table 1.** COSMOGLOBE flagging statistics for each DA. The second column indicates the fraction of data that are removed by the official WMAP flags, while the third column indicates the fraction that is additionally discarded in the current processing for computational reasons. The fourth column indicates the total fraction of data actually used to generate the final maps.

Band	Flagged (%)	Discarded (%)	Used (%)
K	1.72	0.87	97.4
Ka	1.64	0.88	97.5
Q1	1.84	0.84	96.5
Q2	1.62	0.81	97.6
V1	1.62	1.10	97.3
V2	1.61	1.01	97.4
W1	1.76	1.03	97.2
W2	1.60	0.81	97.6
W3	1.61	0.87	97.5
W4	1.60	0.81	97.6

$\mathbf{n} = \mathbf{n}^{\text{w}} + \mathbf{n}^{\text{corr}}$ , a white noise term and a correlated noise term. By definition, the white noise does not have any correlations between adjacent pixels, so that any pixel-pixel covariance should be fully described by realizations of the  $\mathbf{n}^{\text{corr}}$  timestream.

Commander estimates the baseline using the full estimate of the current sky model,  $\mathbf{r} = \mathbf{d} - \mathbf{gs}^{\text{tot}} = \mathbf{b} + \mathbf{n}$ . Modeling  $\mathbf{b} = b_0 + b_1 \Delta t$ , we solve for  $b_0$  and  $b_1$  using linear regression in each timestream while masking out samples that lie within the processing mask. Strictly speaking, this is breaking the Gibbs chain, as we are not formally sampling  $b_0$  and  $b_1$  for each TOD chunk. In practice, baseline estimation uncertainty propagates to correlated noise realizations and PSD parameters, as discussed below.

The approach detailed by Hinshaw et al. (2003) and the Commander implementation differ mainly in two ways. First, the assumed stable timescales are different – the initial WMAP baseline is estimated over one hour timescales, and assumed to be an actual constant, whereas Commander assumes constant values through the entire time chunk, which is 3–7 days depending on the band in question, but allows a linear term in the baseline. Second, the two methods differ in how they treat nonlinear residuals in the first-order baseline model. As noted by Hinshaw et al. (2003), residual baseline variations manifest as correlated noise stripes in the final maps, and WMAP9 solves this using a time-domain filter, downweighting the data based off of the noise characterization. This is fundamentally similar to the Commander3 approach, which accounts for this as part of the correlated noise component. The main advantages of the latter is that it allows for proper error propagation at all angular scales without the use of a dense pixel-pixel noise covariance, and also that it provides a convenient means for inspecting the residuals visually by binning the correlated noise into a sky map.

## 3. Data and data processing

We describe the delivered WMAP data in Sect. 3.1, then describe the treatment we apply to make them compatible with Commander3 in Sect. 3.2. Finally, we describe the computational requirements in Sect. 3.3.

**Table 2.** Computational resources required for end-to-end COSMOGLOBE processing. All times correspond to CPU hours, and all data volumes are reported in GB. Reported times are averaged over more than 100 samples, and vary by  $\lesssim 5\%$  from sample to sample. Note that the average cost per sample takes into account the undersampling of 70, V, and W.

ITEM	30	44	70	K	Ka	Q1	Q2	V1	V2	W1	W2	W3	W4	SUM	
<i>Data volume</i>															
Compressed TOD volume .....	86	178	597	13	12	15	15	19	18	26	26	26	26	1 053	
<i>Processing time (cost per run)</i>															
TOD initialization/IO time .....	1.8	2.5	7.8	0.7	0.6	0.8	0.7	0.9	0.8	1.3	1.3	1.0	0.9	21.1	
Other initialization .....														14.6	
<b>Total initialization .....</b>														<b>35.7</b>	
<i>Gibbs sampling steps (cost per sample)</i>															
Huffman decompression .....	1.2	2.2	23.2	0.8	0.9	1.1	1.1	1.5	1.4	2.0	2.0	2.0	2.0	41.4	
Array allocation .....	0.4	0.9	51.6	1.3	1.3	1.5	1.5	3.1	3.3	4.0	3.8	4.0	4.0	80.7	
TOD projection ( $P$ operation) .....	0.9	2.0	12.3	6.1	7.1	8.7	8.9	11.4	11.3	15.9	15.8	15.7	15.8	131.9	
Sidelobe evaluation .....	1.2	2.6	9.5	3.0	3.5	4.1	4.2	5.5	5.4	7.8	7.7	7.7	7.5	69.7	
Orbital dipole .....	0.9	2.0	9.0	1.2	1.5	1.8	1.9	2.6	2.5	3.8	3.8	3.8	3.8	38.6	
Gain sampling .....	0.6	0.9	2.2	1.3	1.3	0.8	0.8	1.3	1.3	1.2	1.2	1.2	1.2	15.3	
1 Hz spike sampling .....	0.3	0.4	1.9											2.7	
Correlated noise sampling .....	2.1	4.3	24.8	2.7	2.9	3.7	3.8	6.2	5.4	7.7	7.4	6.9	8.3	86.4	
Correlated noise PSD sampling .....	5.0	6.2	1.6	0.3	0.3	0.3	0.3	0.5	0.5	0.7	0.6	0.6	0.7	17.6	
TOD binning ( $P^t$ operation) .....	0.1	0.1	10.5	0.8	0.8	1.0	1.0	1.7	1.6	2.4	2.4	2.4	2.4	27.2	
Mapmaking .....						9.2	9.7	13.1	12.7	21.7	20.2	35.4	34.9	36.1	39.3
MPI load-balancing .....	1.2	1.7	9.2	2.2	2.0	2.2	2.1	3.6	3.3	4.8	4.6	4.5	4.6	46.0	
Sum of other TOD processing .....	0.7	1.6	13.1	0.1	0.2	0.5	0.4	0.7	0.8	0.9	1.0	0.9	1.2	22.1	
<b>TOD processing cost per sample</b>	<b>14.6</b>	<b>24.9</b>	<b>169.7</b>	<b>28.8</b>	<b>31.5</b>	<b>38.7</b>	<b>38.7</b>	<b>59.8</b>	<b>57.0</b>	<b>86.6</b>	<b>85.2</b>	<b>85.8</b>	<b>90.8</b>	<b>812.1</b>	
Amplitude sampling .....														16.2	
Spectral index sampling .....														32.1	
<b>Average cost per sample .....</b>														<b>418.9</b>	

### 3.1. Publicly available WMAP products

The full *WMAP* dataset is hosted at the Legacy Archive for Microwave Background Data Analysis (LAMBDA).<sup>4</sup> In addition to the primary scientific products, e.g., cosmological parameters, CMB power spectra and anisotropy maps and frequency maps, the time-ordered data (TOD) can be downloaded, both in uncalibrated and calibrated form.<sup>5</sup> In principle, thanks to these data and the explanatory supplements (Greason et al. 2012), the entire data analysis pipeline can be reproduced from TOD in digital units (du) to frequency maps.

For this analysis, we keep certain instrumental parameters fixed to the reported values. For example, we have made no attempts to rederive the pointing solutions, re-estimate the main beam response and far sidelobe pickup, or recover data that were flagged in the *WMAP* event log. These and other analyses, such as estimating the bandpass shift over the course of the mission, are certainly possible within the larger Gibbs sampling framework. However, in this work we limit ourselves to recalibrating the TOD, estimating the noise properties, and applying bandpass corrections to the data before mapmaking.

### 3.2. TOD preprocessing and data selection

The full nine-year *WMAP* archive spans from 10 August 2001 to 10 August 2010, with the raw uncalibrated data comprising 626 GB. A little over 1 % of the data were lost or rejected due to incomplete satellite telemetry, thermal disturbances, spacecraft

anomalies, and station-keeping maneuvers, with an extra 0.1 % rejected due to planet flagging (Bennett et al. 2003b; Hinshaw et al. 2007, 2009; Bennett et al. 2013). The final results reported by Bennett et al. (2013) included roughly 98.4 % of the total data volume. A full accounting of all data cuts can be found in Table 1.8 of Greason et al. (2012). In this analysis we flag the same data indicated in the fiducial *WMAP* analysis, and use the same planet flags.

As shown by Galloway et al. (2022a), a large fraction of Commander3’s computational time is spent performing Fast Fourier Transforms (FFTs) on individual scans. Rather than truncating datastreams to have lengths equal to “magic numbers” for which FFTW (Frigo & Johnson 2005) is fastest, as was done in the BEYONDPLANCK analysis, we redistribute the data into scans of length  $2^N$ , where  $N = 22$  for K–Q,  $N = 23$  for V–W. This yields scans with lengths of 6.21 days for K- and Ka-band, 4.97 days for Q-band, 7.46 days for V-band, and 4.97 days for W-band.<sup>6</sup> These datastream lengths are short enough to be processed quickly and distributed efficiently across multiple processors, while being long enough to properly characterize the noise properties of the timestreams, whose  $f_{\text{knee}}$  values are on the order 1 mHz. Most importantly, FFTW performs fastest when the datastream is of length  $2^N$ .

When redistributing the data, timestreams of length  $2^N$  were interrupted by events logged in Table 1.8 of Greason et al. (2012). When we encountered these events, interrupted TOD segments were appended to the previous TOD, in most cases creating TODs with lengths  $> 2^N$ . We found that events of length  $< 2^N$  were too short to accurately estimate the noise PSD param-

<sup>4</sup> [https://lambda.gsfc.nasa.gov/product/wmap/dr5/m\\_products.html](https://lambda.gsfc.nasa.gov/product/wmap/dr5/m_products.html)

<sup>5</sup> [https://lambda.gsfc.nasa.gov/product/wmap/dr5/tod\\_info.html](https://lambda.gsfc.nasa.gov/product/wmap/dr5/tod_info.html)

<sup>6</sup> Note that scans with equal  $n_{\text{TOD}}$  cover different lengths of time due to the different sampling rate for each frequency.

eters. This criterion led us to discard these otherwise useful data. In addition, when  $> 10\%$  of the TOD are flagged, the large number of gaps in the data makes the constrained realizations computationally more expensive. Given that data near many large gaps are more likely to have unmodeled effects than stable data, and they are more expensive to process, we chose to remove these from the analysis. Together, these two effects led to  $\simeq 1\%$  of the data to be discarded. We summarize the full flagging statistics for our maps in Table 1. In total, the COSMOGLOBE maps use about 1 % less data than the WMAP9 official products. The total difference in data volume can be entirely accounted for by the cuts described in this paragraph.

### 3.3. Computational resources and future plans

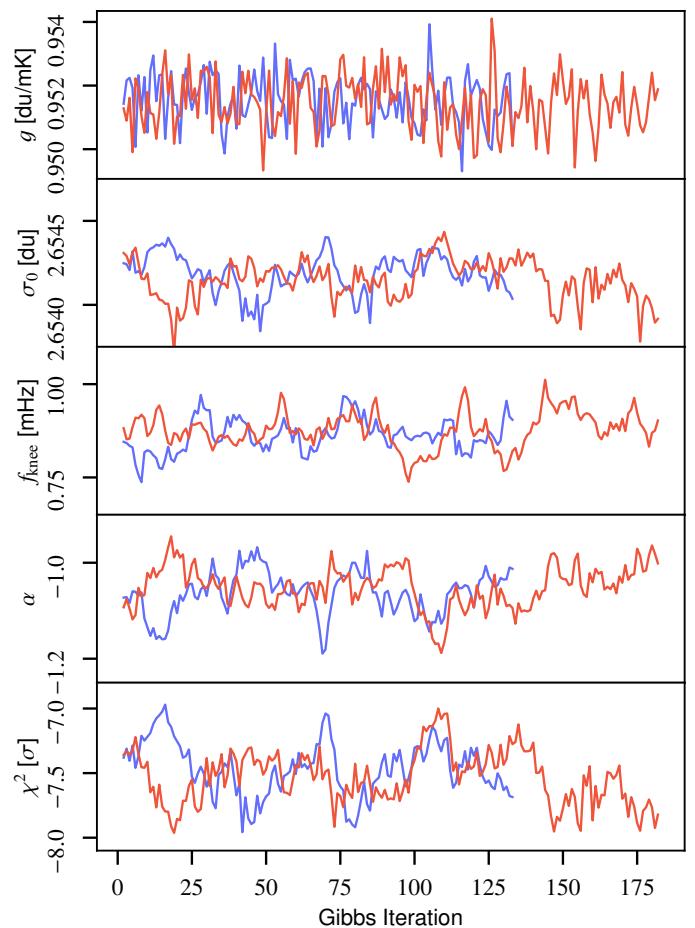
A key motivation of the current analysis is to evaluate whether it is feasible to perform a joint analysis of two datasets simultaneously, each with its own particular processing requirements and algorithmic treatment. One of the results from Watts et al. (2022) was that most of the data processing procedures for WMAP and *Planck* LFI overlapped, with the notable exception of mapmaking. While the algorithmic requirements have been discussed in Sect. 2, we have not yet quantified the requirements in terms of RAM and CPU hours. In Table 2, we enumerate the RAM requirements and CPU time for each sampling step using a single AMD EPYC 7H12, 2.6 GHz cluster node with 128 cores and 2 TB of memory. As such, wall runtimes can be obtained by dividing all numbers in Table 2 by 128.

Despite the relatively small data volume spanned by WMAP, the CPU time is comparable to each of the LFI channels. The single largest reason for this is the mapmaking step, which requires looping over the entire dataset for each matrix multiplication, a process which must be repeated  $\sim 20$  times. As discussed in Sec. 2.7.2, this is vastly sped up by the use of a low resolution preconditioner, reducing the number of iterations by an order of magnitude.

Additionally, operations that require creating timestreams for each detector, i.e., TOD projection, sidelobe evaluation, and orbital dipole projection, take much longer than expected from a pure data volume scaling. Part of this is due to each WMAP radiometer needing to evaluate the sky in two pixels simultaneously, doubling the expected workload, but the other issue is that we are unable to benefit from the ring-clustering based TOD distribution scheme used for LFI. Due to WMAP’s more complex scan strategy and detector geometry, it is impossible to cluster scans with similar pixel coverage onto a single core, which makes pixel-space lookup operations less efficient in this case.

Gain sampling and correlated noise sampling include multiple FFTs. Typical LFI TODs are of length  $\sim 200\,000$ , an order of magnitude smaller than the WMAP TODs of length  $\sim 5\,000\,000$ . Despite the TOD lengths being pre-determined to be  $2^N$ , this extra length still results in longer run times for equivalent data volumes, but does yield noise information on much longer time scales than we have for LFI.

For the current analysis, which aims primarily to derive posterior-based WMAP frequency maps, we produce a total of 500 main Gibbs samples, divided into two chains. Noting that the computational cost of the *W*-channel carries almost half of the total expense of the WMAP TOD processing, while being of less scientific importance than, say, the *K*-band, we choose to only re-process this channel every fourth main sample. Likewise, we only reprocess the *V*-band every other main sample, and the LFI 70 GHz sample every fourth sample. The total cost for producing 500 WMAP *K*, *Ka*, *Q*, *Planck* 30, and 44 GHz samples,

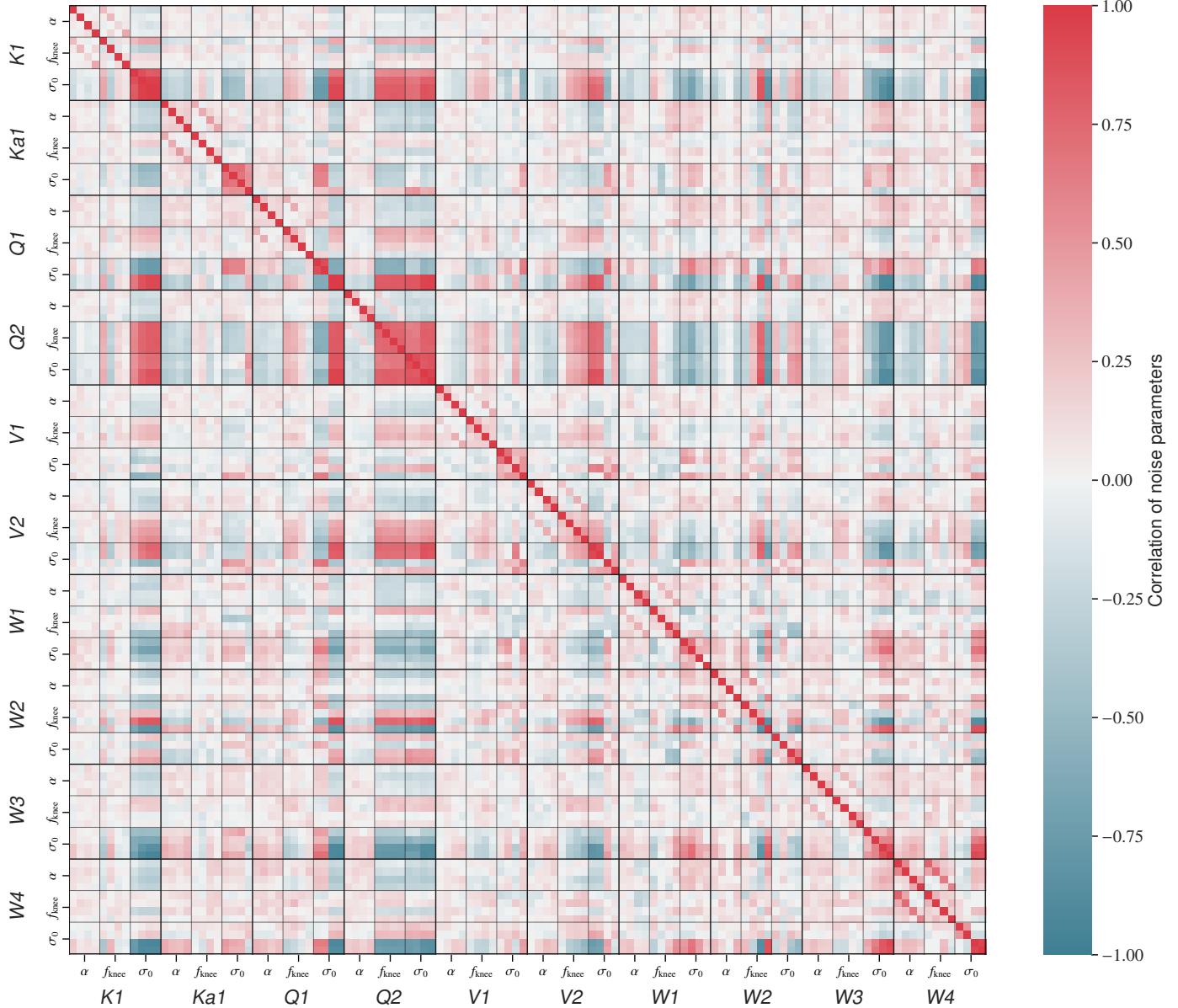


**Fig. 3.** Trace plots of the *K*113 gain and noise parameters for a single scan starting on MJD 52285.2. The two colors correspond to the two independent Markov chains produced in this analysis.

250 *V*-band samples, and 125 *W*-band and 70 GHz samples is 210k CPU-hrs, and the total walltime is 33 days. Noting that the BEYONDPLANCK analysis required 4000 samples to reach full convergence in terms of the optical depth of reionization (Paradiso et al. 2022), a corresponding complete LFI+WMAP analysis will cost about 1.7M CPU-hrs, and take about 9 months of continuous runtime on two cluster nodes. While entirely feasible, this is sufficiently expensive that we choose to perform the analysis in two stages; first we present preliminary frequency maps in the current paper, and use these to identify potential outstanding issues, either in terms of data model or Markov chain stability. An important goal of this phase is also to invite the larger community to study these preliminary maps, and thereby identify additional problems that we may have missed. Then, when all issues appear to have been resolved, we will restart the process, and generate sufficient samples to achieve full convergence.

## 4. Instrumental parameters

We are now ready to present the main results from the COSMOGLOBE DR1 analysis, which may be summarized in terms of the joint posterior distribution. For organizational purposes, we will discuss instrumental parameters, frequency maps, and astrophysical results separately in this and the following two sections, but it is important to remember that these results are all derived from one single highly multivariate posterior distribution, and every parameter is in principle correlated with all others. In this



**Fig. 4.** Noise parameter correlation matrix. We average over all Gibbs samples of the noise parameters  $\xi^n = \{\alpha, f_{\text{knee}}, \sigma_0\}$  for each PID. We then find the correlation in time between these averages for the different bands and detector. The results here are for the calibrated white noise level,  $\sigma_0[\text{mK}]$ . The values for each detector are ordered 13, 14, 23, and 24.

section, we focus on instrumental parameters, starting with visual inspection of the basic Markov chains and posterior means, before considering each instrumental parameter in turn.

#### 4.1. Markov chains, correlations and posterior mean statistics

To build intuition regarding the general Markov chain properties, we show in Fig. 3 the Markov chains for the gain and noise parameters for one randomly selected diode (*K113*) and scan. Each panel corresponds to one single parameter, and the observed variation quantify the uncertainty in that single parameter due to the combination of white noise and correlations with other parameters. Here we immediately see that the different parameters have quite different correlation lengths; the gain (shown in the top panel) has a very short auto-correlation length, as in just a few samples, while the noise parameters have typical correla-

tion lengths of a few tens of samples. Even for these parameters, however, the full set of 500 samples provides a fairly robust estimate of the full marginal mean and uncertainty.

The bottom panel shows the reduced normalized  $\chi^2$  for the same scan in units of  $\sigma$ , and we see that this also shows similar correlation lengths as the noise parameters. This makes intuitive sense since the TOD residual at the level of a single-sample is strongly noise dominated. In contrast, small variations in either the sky signal or gain have relatively small impacts on this particular  $\chi^2$ ; the goodness of fit of such global parameters is better measured through map-level residuals and  $\chi^2$ 's. In this respect, we also note that the absolute value of the TOD-level  $\chi^2$  is for this particular scan about  $-7.5\sigma$ , which at first sight appears as a major goodness of fit failure. However, it is important to recall that a typically scan contains about five million data points, and this statistic is therefore extremely sensitive to any deviation in the noise model. Specifically, the reduced  $\chi^2$  for this

particular scan is  $\chi^2_{\text{raw}}/n_{\text{tot}} = 0.993$ , which corresponds to an over-estimation of the white noise level of only 0.3 %. Furthermore, as discussed in Sect. 2.7.1, we currently assume a strict  $1/f$  noise model for the *WMAP* noise, while the true *WMAP* noise is known to exhibit a very slight non-white noise excess at high frequencies (Watts et al. 2022). Properly modelling such non-white high-frequency noise is therefore an important goal for the next COSMOGLOBE data release. Such work is also a vital step in preparing for integration of other types of experiments with non-white noise into the framework, such as *Planck* HFI. However, in absolute terms, the impact of this model failure is very limited, and not likely to significantly affect any astrophysical results; it is primarily a limitation for TOD-level goodness of fit testing.

Going one step higher in complexity, Fig. 4 shows pairwise correlations between the various noise parameters for all DAs, averaged over all Gibbs samples and scans. It is important to note that a non-zero correlation in this plot does not indicate the that specific noise realization is correlated between DAs, but only that the noise PSD parameters are correlated. This is expected due to the *WMAP* satellite motion around the Sun, which induces an annual variation in the system temperature. This correlation plot therefore primarily quantifies the sensitivity to this common-mode signal for each radiometer. Most notably, we see that the  $Q2$  radiometer exhibits particularly strong correlations, and we also note that  $\sigma_0$  is generally more susceptible to these variations than  $f_{\text{knee}}$  and  $\alpha$ .

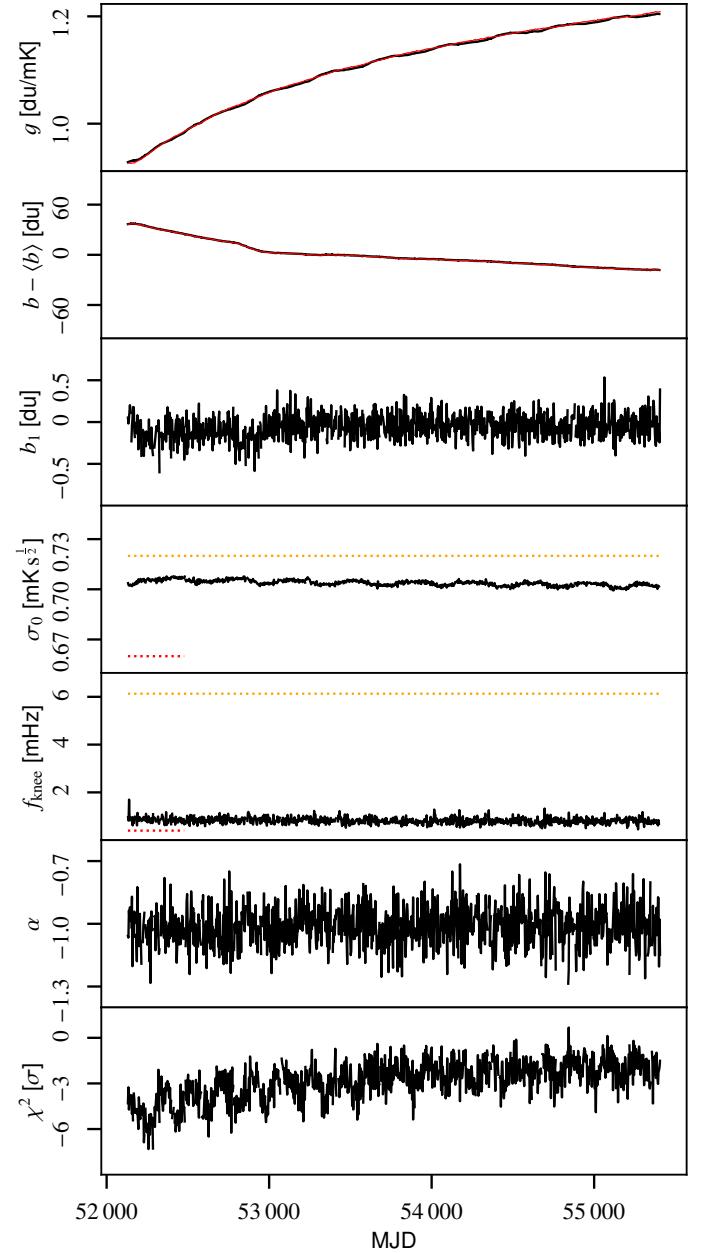
Next, in Fig. 5 we show posterior mean values for each instrumental parameter for the same  $K113$  diode, in this case plotted as a function of time throughout the entire mission. The panels show, from top to bottom, 1) gain; 2) the difference between the baseline mean and its full-mission average; 3) the baseline slope; 4) the white noise level; 5) the correlated noise knee frequency; 6) the correlated noise slope; and 7) the TOD-level  $\chi^2$ . The COSMOGLOBE results are shown as black curves, while the *WMAP* results are (for the gain and baseline) shown as red curves; dotted red and orange line corresponds to the first-year *WMAP* and Goddard Space Flight Center (GSFC) laboratory measurements, respectively. For brevity, we have only shown the results for one single diode here. However, a complete survey of all instrumental parameter posterior means for all 40 diodes is provided in Appendix A, and all individual samples are also available in a digital format as part of the COSMOGLOBE DR1.

#### 4.2. Gain and baselines

We now consider the gain and baseline parameters in greater detail, and aim to compare our estimates with the *WMAP9* products. Unfortunately, the *WMAP9* gain and baseline estimates are not directly available in terms of easily accessible and public data products, but only in terms of the general parametric models. For instance, the *WMAP* gain model reads Greason et al. (2012)

$$g = \alpha \frac{\bar{V} - V_\circ - \beta(T_{\text{RXB}} - 290 \text{ K})}{T_{\text{FPA}} - T_\circ} + (m\Delta t + c), \quad (48)$$

where  $\bar{V}$  represents the radio frequency bias powers per detector;  $T_{\text{RXB}}$  and  $T_{\text{FPA}}$  are the receiver box and focal plane assembly temperatures, which are recorded every 23.04 s;  $\alpha$ ,  $V_\circ$ ,  $\beta$ ,  $T_\circ$ ,  $m$ , and  $c$  are all free parameters that are fit to a constant value across the mission for each radiometer. Evaluating this model as a function of  $T_{\text{RXB}}$  and  $T_{\text{FPA}}$  requires the housekeeping data

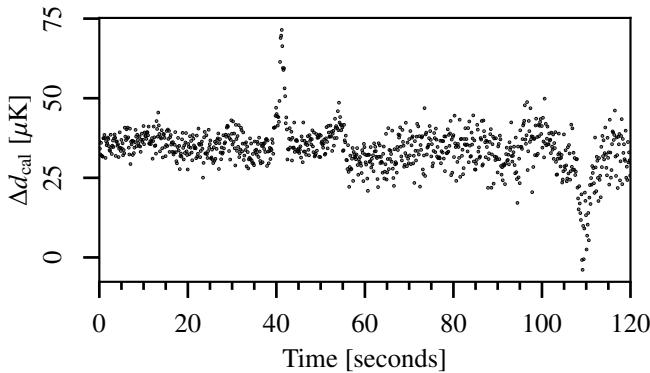


**Fig. 5.** Overview of  $K113$ . The red solid lines in first and second panel are the regressed gain and baseline from *WMAP9*, while the black lines in all panels are samples from the COSMOGLOBE Gibbs chain. The red dashed and yellow dashed lines are reported  $\sigma_0$  and  $f_{\text{knee}}$  values from the first-year *WMAP* data analysis and GSFC measurements, respectively.

for the thermistor that was physically closest to the relevant radiometer's focal plane on the satellite. The free parameters are fully tabulated in the *WMAP* Explanatory Supplement (Greasen et al. 2012), but the physical layout of the thermistors in the focal plane is not readily available. We therefore do not attempt to reproduce the gain model given in Eq. (48).

Rather, we estimate the gains and baselines by comparing the uncalibrated *WMAP* data with the calibrated *WMAP* data, after subtracting a far sidelobe contribution convolved with the delivered *WMAP9* DA maps plus the Solar dipole. We find that the calibrated and uncalibrated data can be related by

$$d_t^{\text{raw}} = g(d_t^{\text{cal}} + s_t^{\text{sl}}) + \sum_{i=0}^3 c_i(t - t_0)^i, \quad (49)$$



**Fig. 6.** Difference between the COSMOGLOBE  $d_{\text{cal}} = d/g - b - s_{\text{sl}}$  and the delivered calibrated TOD from WMAP.

where the second term is a cubic polynomial with coefficients  $c_i$  referenced to the time at the beginning of the scan  $t_0$ . The red curves in the top two panels of Fig. 5 correspond to these estimates. At least visually, the agreement between the WMAP and COSMOGLOBE gain and baseline models appear reasonable at this level and for this diode.

A complete comparison between the WMAP and COSMOGLOBE gain and baseline models for all diodes is provided in Appendix A. In particular, Fig. A.1 shows the baseline differences as a function of time, and here we see that most diode differences scatter around a constant value that is close to zero; the precise constant value is of limited importance, since that only corresponds to a difference in the overall monopole of the maps, which for WMAP is determined through post-processing. However, there are a few notable features. First, we see that the two Q11 diodes exhibit large variations at the very beginning of the mission, with typical values of a few  $\mu\text{K}$ 's, and individual scans show notable spikes for many diodes. These are all relatively isolated in time, and will therefore have relatively minor impact on the final maps. Far more significant are the  $W$ -band differences, for which one sees both slow drifts and abrupt changes. Furthermore, in many cases they vary notably between diodes within the same DA, and this will obviously translate into differences in the large-scale polarization maps derived from the two pipelines.

Similarly, Fig. A.3 compares the gain solutions directly, while Fig. A.4 shows the fractional differences in units of percent. Overall, we see that the two gain models agree to typically about 0.5 % in an absolute sense, and better than typically 0.1 % in terms of relative agreement between neighboring scans. By far the most striking feature in this plot is an annual variation that traces the WMAP satellite's motion around the Sun. In general, such an oscillatory gain behaviour is entirely expected, because of known temperature variations in the satellite. However, the difficulty lies in estimating the magnitude of the oscillations, as different radiometers can respond differently to these temperature variations. In this respect, it is useful to recall that the WMAP and COSMOGLOBE gain estimation algorithms differ at a fundamental level; while the WMAP analysis considers each DA in isolation, and attempts to fit a handful of instrumental parameters, defined by Eq. (48), to the orbital dipole seen by each DA. The COSMOGLOBE analysis considers the problem globally, and attempts to fit all gain parameters to the full sky signal (including both the Solar and orbital CMB dipole) simultaneously, without the use of a strong instrumental model prior. Returning to the absolute gains shown in Fig. A.3, it is difficult to determine visually which approach is better at this level alone, as the

**Table 3.** Transmission imbalance parameters for each WMAP radiometer as estimated in the current analysis (*second column*) and in the official 9-year WMAP analysis (*third column*). Our uncertainties indicate  $1\sigma$  marginal posterior standard deviations.

RADIOMETER	$x_{\text{im}}^{\text{CG}}$	$x_{\text{im}}^{\text{WMAP}}$
K11 . . . . .	$0.00018 \pm 0.00013$	$-0.00067 \pm 0.00017$
K12 . . . . .	$0.00388 \pm 0.00015$	$0.00536 \pm 0.00014$
Ka11 . . . . .	$0.00339 \pm 0.00012$	$0.00353 \pm 0.00017$
Ka12 . . . . .	$0.00150 \pm 0.00010$	$0.00154 \pm 0.00008$
Q11 . . . . .	$0.00081 \pm 0.00016$	$-0.00013 \pm 0.00046$
Q12 . . . . .	$0.00517 \pm 0.00027$	$0.00414 \pm 0.00025$
Q21 . . . . .	$0.00985 \pm 0.00042$	$0.00756 \pm 0.00052$
Q22 . . . . .	$0.01235 \pm 0.00011$	$0.00986 \pm 0.00115$
V11 . . . . .	$0.00012 \pm 0.00041$	$0.00053 \pm 0.00020$
V12 . . . . .	$0.00212 \pm 0.00089$	$0.00250 \pm 0.00057$
V21 . . . . .	$0.00246 \pm 0.00012$	$0.00352 \pm 0.00033$
V22 . . . . .	$0.00323 \pm 0.00070$	$0.00245 \pm 0.00098$
W11 . . . . .	$0.01169 \pm 0.00105$	$0.01134 \pm 0.00199$
W12 . . . . .	$0.00442 \pm 0.00109$	$0.00173 \pm 0.00036$
W21 . . . . .	$0.01595 \pm 0.00052$	$0.01017 \pm 0.00216$
W22 . . . . .	$0.01540 \pm 0.00167$	$0.01142 \pm 0.00121$
W31 . . . . .	$-0.00089 \pm 0.00010$	$-0.00122 \pm 0.00062$
W32 . . . . .	$0.00354 \pm 0.00084$	$0.00463 \pm 0.00041$
W41 . . . . .	$0.02734 \pm 0.00219$	$0.02311 \pm 0.00380$
W42 . . . . .	$0.01882 \pm 0.00282$	$0.02054 \pm 0.00202$

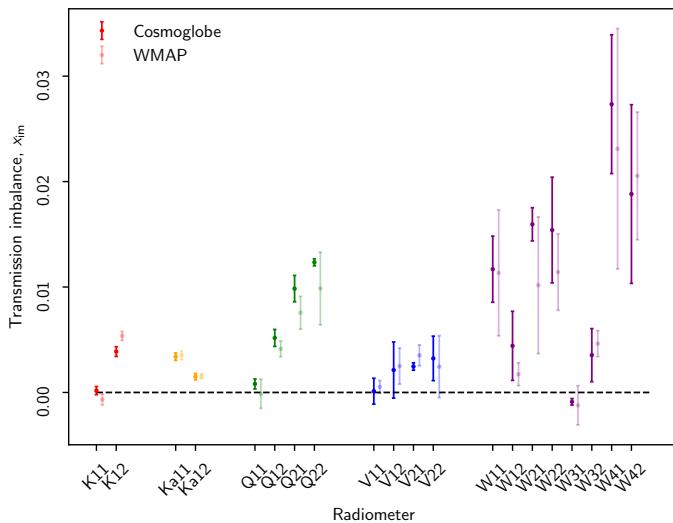
two models are quite similar; in some cases, such as  $Ka124$  and  $Q214$ , the WMAP model oscillates more strongly than the COSMOGLOBE model, while in others, such as  $K113$  and  $K114$ , the opposite is true. We also see the impact of the strong instrumental priors in the WMAP solution particularly well in  $W$ -band, where the COSMOGLOBE gains are far more noisy than the WMAP gains.

The impact of these differences at the TOD level is illustrated in Fig. 6, which shows the calibrated COSMOGLOBE timestream  $d/g - s_{\text{sl}}$  minus the WMAP calibrated signal in units of microkelvin. The most prominent feature is a  $\sim 25 \mu\text{K}$  offset, which is unsurprising, given the different treatment of baselines in our two pipelines. The second obvious difference is a series of spikes associated with Galactic plane crossings. Differences of order  $50 \mu\text{K}$  are seen where the absolute sky brightness is about  $10 \text{ mK}$ , and this is equivalent to  $\sim 0.5 \%$  deviations in the gain solution. This is twice as large as the 0.2 % uncertainty estimated in Bennett et al. (2013) based on end-to-end simulations.

Another interesting feature in Fig. 6 is slow correlated variations at a timescale of  $\sim 20$  sec timescale. There is nothing in the COSMOGLOBE instrument model that varies on such short timescales, and this must therefore come from WMAP. The most likely explanation is the fact that the the WMAP gain model depends directly on housekeeping data that are recorded with a 23.04 sec sample rate, and these values appear to have been applied without any smoothing, resulting in sharp jumps in the final WMAP gain model. At the same time, it is also important to note that the COSMOGLOBE gain model does not include any time-varying structure within a single scan, and if any artifacts resulting from this are identified in the current products, it may be worth incorporating housekeeping data in a future COSMOGLOBE data release.

#### 4.3. Transmission imbalance

Closely related to the gain model is the transmission imbalance factor,  $x_{\text{im}}$ , quantifying the difference between responsivity in the two horns, as described in Sect. 2.2. These are listed for



**Fig. 7.** Comparison of the transmission imbalance factors,  $x_{\text{im}}$ , estimated by COSMOGLOBE (dark colors) and WMAP9 (light colors) for each radiometer.

each radiometer in Table 3 for both COSMOGLOBE and WMAP9; for COSMOGLOBE the reported values correspond to marginal posterior means and standard deviations. The same information is plotted in Fig. 7.

At first glance, the two estimates appear in reasonable agreement, and except for  $K$ -band, they are always within  $1\sigma$  from each other. However, there are subtle differences that could play an important role in the final maps and uncertainty estimates. The reason for this is that these parameters couple directly to the astrophysical sky signal, and in particular to the bright 3 mK Solar CMB dipole. Even an inaccuracy at the  $O(10^{-3})$  level can therefore in principle excite correlated large scale artifacts at the microkelvin level, which is comparable to, or larger than, the expected cosmological reionization of about  $0.5\mu\text{K}$ .

With this in mind, we first note that the  $K$ -band imbalance factors differ by  $2\text{--}3\sigma$ . For this particular channel, it is important to note both that the foregrounds are very bright and that the effective bandpass of each radiometer differ. Minor details in both foreground and bandpass modelling are therefore likely to have a significant impact of the estimated parameters.

Second, we see that for many several channels, the two methods disagree strongly on the marginal uncertainties. For instance, the WMAP  $Q22$  uncertainty is about a factor of ten larger than the COSMOGLOBE uncertainty, and the corresponding mean differs from the COSMOGLOBE mean by about  $8\sigma$  as measured in units of COSMOGLOBE uncertainties. These differences are important both in terms of final sky maps (which depend on the mean value) and error propagation. In particular, it is important to note that while the COSMOGLOBE sampling approach accounts for all couplings between the specific value of  $x_{\text{im}}$  and all other parameters (gain, baselines, correlated noise, CMB dipole, large scale polarization, etc.) at every single step of the Markov chain, the WMAP approach only marginalizes over two linear templates in the low-resolution covariance matrix. These two templates are derived by changing  $x_{\text{im}}$  by respectively 10 and 20 % with respect to their mean values, respectively, and recomputing the maps. This linear low resolution approach can only capture a limited subvolume of the full nonlinear effect of transmission imbalance uncertainties. Even cases for which the mean estimates formally agree within  $1\sigma$  may therefore in practice result in significantly different sky maps. We will return to this issue in Sect. 5.

#### 4.4. Instrumental noise

Next, we consider the instrumental noise parameters,  $\xi_n = \{\sigma_0, f_k, \alpha\}$ . In this case, we recall three major differences between the COSMOGLOBE and WMAP analysis. First, while we model the noise explicitly with a  $1/f$  noise profile in Fourier domain, the WMAP analysis adopts a model independent approach by simply measuring the autocorrelation function directly. A notable advantage of the latter approach is that it naturally accounts for the non-white noise at high frequency without algorithmic modifications, while this has to be added manually in the parametric COSMOGLOBE approach. A second difference is the fact that while WMAP uses 1- or 24-hour segments to estimate the noise model, we use 3–5 days, and are therefore able to trace noise correlations to longer timescales. Thirdly, while WMAP assumed the noise filter to be constant within each year of operations, we allow it to vary between scans, that is, on a timescale of days.

With these differences in mind, Figs. A.5–A.7 provides a complete overview of the noise parameters for all 40 WMAP diodes. As in Fig. 5, the solid black lines show COSMOGLOBE results, while the dotted red and orange lines show the corresponding 1-year and GSFC measurements (where available). Starting with the white noise level, we see that these are overall relatively constant in time, although with slight traces for annual variations in some channels (e.g.,  $K113$ ); slight instabilities near the beginning and/or end of the mission in other channels (e.g.,  $Ka$ ); and slight drifts in yet others (e.g.,  $Q12$  and  $W32$ ).

When comparing the COSMOGLOBE values with the WMAP values, it is worth noting that WMAP only published results for each diode-pair, not for individual diodes. All WMAP values are therefore the same for each diode pair. Still, from the COSMOGLOBE results, which are reported individually for each diode, we see that diode pairs generally have quite similar white noise levels and vary at most by a percent.

To facilitate a more quantitative comparison, Table 4 compares the COSMOGLOBE posterior mean results with the reported WMAP results. Note that for  $\sigma_0$ , the COSMOGLOBE values have been scaled by a factor of  $\sqrt{2}$ , in order to account for the fact that these apply to individual diodes, as opposed to diode-pairs. Both in Table 4 and Fig. A.5, we see that about half of the COSMOGLOBE values lie between the two WMAP results, while the other half are higher. In particular the  $W$ -band noise levels are much higher in the COSMOGLOBE solution, sometimes by as much as 50 %.

In this respect, it is worth recalling from Sect. 2.7.1 that the white noise level in raw  $du$  is in COSMOGLOBE not strictly sampled from the full posterior distribution, but rather estimated deterministically from the highest frequencies. This makes our estimate more sensitive to possible colored noise at high frequencies (Watts et al. 2022). At the same time, the calibrated white noise level  $\sigma_0[V] = \sigma_0[du]/g$  depends on the gain, and this allows us to test the effects of the calibration on the instrument sensitivity itself. The calibrated white noise level follows a biannual trend indicative of a system temperature variation, which is to be expected given the radiometer equation

$$\sigma_0[V] \propto gT_{\text{sys}}. \quad (50)$$

Aside from an overall amplitude shift due to the absolute calibration variation, the shape of the white noise level is stable throughout the Gibbs chain.

Another issue worth pointing out is the fact that we are not yet accounting for correlations between the white noise in diode pairs. However, these are write something more here.

**Table 4.** Summary of noise properties.

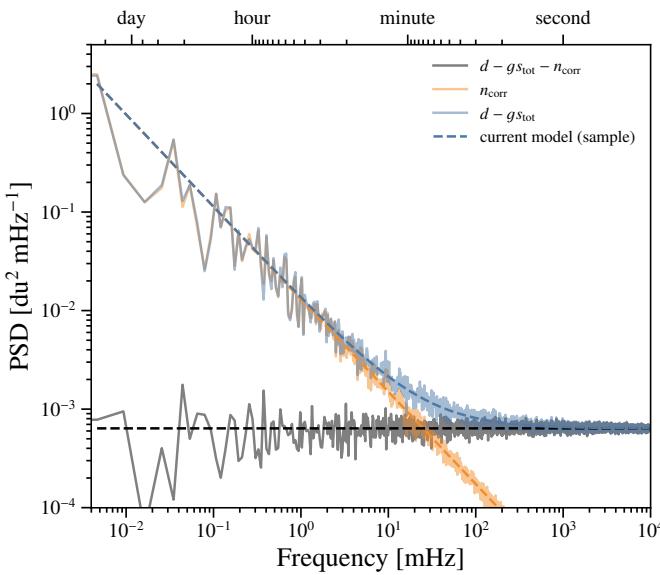
Radiometer ..	Diode	Sensitivity, $\sigma_0$ (mK $\sqrt{s}$ )			Knee frequency, $f_{\text{knee}}$ (mHz)			Slope, $\alpha$
		GSFC	WMAP	$CG/\sqrt{2}$	GSFC	WMAP	$CG/\sqrt{2}$	
K11 .....	1	0.72	0.66	$0.704 \pm 0.002$	6.13	0.4	$0.82 \pm 0.20$	$-1.01 \pm 0.10$
	2			$0.708 \pm 0.003$			$0.63 \pm 0.14$	$-0.95 \pm 0.10$
K12 .....	1	0.87	0.75	$0.796 \pm 0.004$	5.37	0.51	$0.42 \pm 0.19$	$-0.93 \pm 0.12$
	2			$0.780 \pm 0.005$			$0.71 \pm 0.15$	$-1.02 \pm 0.10$
Ka11 .....	1	0.75	0.71	$0.788 \pm 0.001$	1.66	0.71	$1.20 \pm 0.22$	$-1.02 \pm 0.09$
	2			$0.777 \pm 0.001$			$1.19 \pm 0.22$	$-1.02 \pm 0.09$
Ka12 .....	1	0.77	0.72	$0.788 \pm 0.003$	1.29	0.32	$0.62 \pm 0.16$	$-0.99 \pm 0.11$
	2			$0.784 \pm 0.001$			$0.63 \pm 0.13$	$-1.01 \pm 0.11$
Q11 .....	1	0.99	0.92	$0.998 \pm 0.002$	3.21	1.09	$1.06 \pm 0.16$	$-1.09 \pm 0.09$
	2			$0.992 \pm 0.002$			$1.06 \pm 0.16$	$-1.10 \pm 0.09$
Q12 .....	1	0.95	1.02	$1.159 \pm 0.007$	3.13	0.35	$0.45 \pm 0.47$	$-0.98 \pm 0.11$
	2			$1.146 \pm 0.007$			$0.83 \pm 0.14$	$-1.00 \pm 0.09$
Q21 .....	1	0.89	0.85	$0.908 \pm 0.002$	1.92	5.76	$2.88 \pm 0.37$	$-1.10 \pm 0.07$
	2			$0.906 \pm 0.002$			$3.22 \pm 0.56$	$-1.10 \pm 0.06$
Q22 .....	1	1.04	0.99	$1.074 \pm 0.004$	4.61	8.62	$3.95 \pm 0.54$	$-1.11 \pm 0.06$
	2			$1.064 \pm 0.003$			$4.05 \pm 0.64$	$-1.11 \pm 0.06$
V11 .....	1	1.25	1.22	$1.551 \pm 0.003$	2.56	0.09	$1.27 \pm 0.15$	$-0.90 \pm 0.06$
	2			$1.539 \pm 0.003$			$1.19 \pm 0.14$	$-0.89 \pm 0.06$
V12 .....	1	1.07	1.11	$1.398 \pm 0.002$	4.49	1.41	$2.11 \pm 0.20$	$-0.97 \pm 0.05$
	2			$1.432 \pm 0.002$			$1.88 \pm 0.17$	$-0.96 \pm 0.05$
V21 .....	1	1.01	0.97	$1.241 \pm 0.298$	2.43	0.88	$1.50 \pm 0.24$	$-0.95 \pm 0.07$
	2			$1.217 \pm 0.294$			$1.60 \pm 0.26$	$-0.97 \pm 0.06$
V22 .....	1	1.13	1.1	$1.443 \pm 0.300$	3.06	8.35	$4.01 \pm 0.85$	$-1.00 \pm 0.08$
	2			$1.415 \pm 0.316$			$3.08 \pm 0.65$	$-1.01 \pm 0.08$
W11 .....	1	1.18	1.35	$1.938 \pm 0.005$	16.2	7.88	$5.59 \pm 0.53$	$-0.94 \pm 0.05$
	2			$1.895 \pm 0.005$			$8.99 \pm 0.85$	$-0.95 \pm 0.04$
W12 .....	1	1.41	1.61	$2.301 \pm 0.005$	15.1	0.66	$3.91 \pm 0.42$	$-0.89 \pm 0.05$
	2			$2.345 \pm 0.006$			$4.81 \pm 0.53$	$-0.89 \pm 0.05$
W21 .....	1	1.38	1.61	$2.225 \pm 0.007$	1.76	9.02	$13.57 \pm 1.47$	$-0.89 \pm 0.03$
	2			$2.292 \pm 0.006$			$5.06 \pm 0.95$	$-0.93 \pm 0.05$
W22 .....	1	1.44	1.72	$2.291 \pm 0.006$	0.77	7.47	$3.02 \pm 0.53$	$-0.98 \pm 0.05$
	2			$2.232 \pm 0.007$			$7.26 \pm 1.05$	$-0.95 \pm 0.04$
W31 .....	1	1.47	1.65	$2.328 \pm 0.005$	1.84	0.93	$1.30 \pm 0.46$	$-0.99 \pm 0.07$
	2			$2.322 \pm 0.006$			$1.97 \pm 0.28$	$-0.98 \pm 0.06$
W32 .....	1	1.69	1.86	$2.707 \pm 0.015$	2.39	0.28	$1.59 \pm 0.29$	$-0.98 \pm 0.07$
	2			$2.579 \pm 0.015$			$1.40 \pm 0.39$	$-1.00 \pm 0.07$
W41 .....	1	1.6	1.71	$2.519 \pm 0.010$	8.46	46.5	$26.81 \pm 1.83$	$-0.92 \pm 0.04$
	2			$2.479 \pm 0.009$			$24.75 \pm 1.63$	$-0.92 \pm 0.04$
W42 .....	1	1.43	1.65	$2.221 \pm 0.017$	5.31	26.0	$16.10 \pm 1.09$	$-0.94 \pm 0.04$
	2			$2.202 \pm 0.015$			$17.11 \pm 1.19$	$-0.94 \pm 0.04$

In summary, we have not yet been able to identify the cause of the major difference in reported white noise levels at  $W$ -band; while we do detect goodness of fit failures of as much as  $5-10\sigma$  for many of these diodes at the TOD level (see Sect. 4.1), such significances correspond to sub-percent errors in the white noise level. For comparison, a white noise misestimation of 50% would translate into an  $800\sigma\chi^2$  failure. This is left to be understood through future work, but we do not expect it to indicate a real failure in either analysis, but it is more likely just a matter of different conventions.

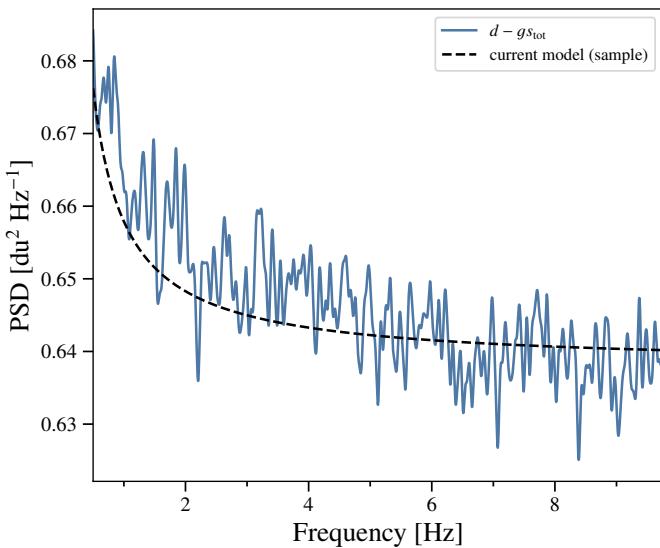
Turning our attention to the low frequency parameters, we see in Table 4 and Fig. A.6 that our knee frequencies lie between the *WMAP* ground and laboratory measurements, almost without exception, which on the one hand indicates generally good agreement between the two analyses. However, on the other hand, it also implies that our values are in fact closer to the *WMAP* laboratory measurements than the *WMAP* flight measurements. This may possibly be due to the longer time-scales used in the COSMOGLOBE analysis.

Most radiometers have constant  $f_{\text{knee}}$  throughout the mission, with a few notable exceptions. First, all  $W$ -band channels display some amount of temporal variation that does not seem to be associated with any sinusoidal features. Second, all  $Q2$  channels,  $V223$ , and  $V224$  all display a similar asymptotic drift in time. We have not found any instrumental effects that share this feature. The PSD slope  $\alpha$  is around  $-1$  for each radiometer, albeit with high scatter for the lower frequencies. As expected, the uncertainty in  $\alpha$  decreases as  $f_{\text{knee}}$  increases, since there are more datapoints to fit below  $f_{\text{knee}}$  where the constraining power on  $\alpha$  is the strongest.

For completeness, Fig. A.8 shows a summary of the reduced normalized  $\chi^2$  for all diodes. The most striking features in these figures are the amplitude and semiannual periodicity. Given the noise model and data residual, we can evaluate the goodness of fit in the form of the relative  $\chi^2$ . Here, we find that approximately half of the radiometers have a  $\chi^2$  value at least  $6\sigma$  above or below the expected value. Given perfect Gaussian residuals, we would expect these values to be within  $\pm 1\sigma$  68% of the



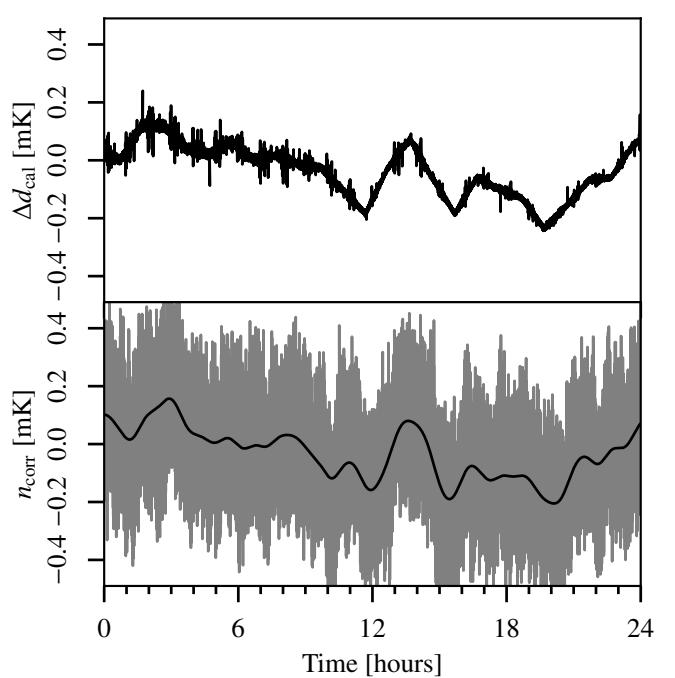
**Fig. 8.** PSD for radiometer W413 that spans MJDs 52252.3–52254.8. The power spectrum of the blue line corresponds to the residual, while the gray line is the residual with a correlated noise realization removed.



**Fig. 9.** PSD for radiometer W413 that spans MJDs 52252.3–52254.8. The black dashed line is a sample of the theoretical PSD, while the blue line is the smoothed residual power spectrum.

time. For a typical  $W$ -band scan of length  $n_{\text{TOD}} = 2^{22}$ , a  $10\sigma$  model failure corresponds to  $\chi^2/n_{\text{TOD}} = 1.003$ . It is therefore exceedingly difficult to look at any given  $WMAP$  scan in the time domain and identify a model failure. To illustrate this, Fig. 8 compares the observed noise PSD with the best-fit model for the W413 diode. This is a  $7\sigma$  outlier; despite this, the  $1/f$  model appears to perform exceedingly well over seven orders of magnitude in frequency.

Only with aggressive smoothing does the model failure become apparent at frequencies 1–10 Hz. This is illustrated in Fig. 9, which shows exactly the same underlying data as in Fig. 8, but heavily smoothed. Here, it is clear that despite fitting the data well at the highest and lowest frequencies, it is in the intermediate range of 1–5 Hz where the  $1/f$  model is a less accurate fit to the residual power spectrum. Part of the cause of

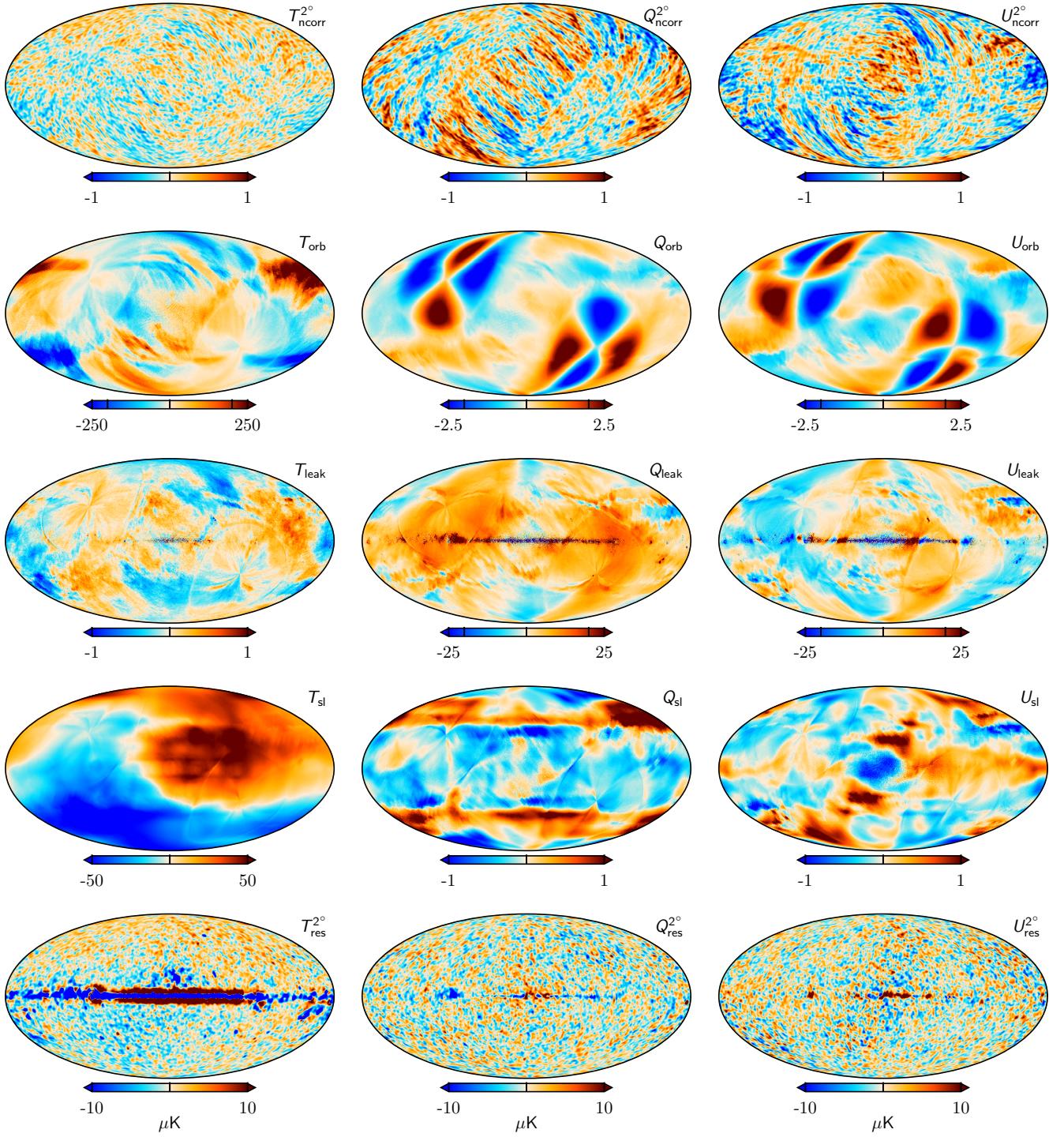


**Fig. 10.** (top): Difference between the COSMOGLobe  $d_{\text{cal}} = d/g - \mathbf{b} - \mathbf{s}_{\text{sl}}$  and the delivered calibrated TOD from  $WMAP$ . (Bottom): Raw correlated noise (gray) and smoothed data with Gaussian kernel (black). This shows more clearly the hourly baseline subtraction from the  $WMAP$  treatment.

this failure is that the white noise level is essentially fixed by the value of the power spectrum at the Nyquist frequency, as it was computed by differencing adjacent samples. The power spectrum has a downward trend beyond above 1 Hz, indicating that the data would be better fit by one or more terms proportional to  $f^\alpha$ . This is phenomenologically similar to the  $WMAP$  collaboration’s approach of describing the time-space autocorrelation as a cubic polynomial in  $\log \Delta t$  (Jarosik et al. 2007).

In practice, the  $1/f$  model has a small effect on the final data products, and was not visible in noise models when we modeled the data in one day scans rather than the longer 3–7 day scans due to the lower  $n_{\text{TOD}}$  giving a higher uncertainty on the relative  $\chi^2$ . Therefore, although this strictly constitutes a deficiency in the model, it is in practice too small to affect the results of the rest of the chain. The downturn of the noise PSD at high frequencies is also present in, e.g., the *Planck* HFI data (Planck Collaboration Int. XLVI 2016, Fig. 1), so improved modeling of this form will be a necessity in future COSMOGLobe endeavors, and will be used to improve the  $WMAP$  data processing.

Before concluding this section, we recall the close relationship between the correlated noise component and the baseline model. This is illustrated in Fig. 10, which shows the difference between the calibrated COSMOGLobe and  $WMAP$  TOD data, i.e., the same as Fig. 6, but plotted for 24 hours instead of 10 minutes. The bottom panel shows the COSMOGLobe correlated noise realization for the same period, both raw and smoothed. The most prominent feature in this figure is a varying signal of amplitude 0.2 mK. This is due to the hourly baseline subtraction mentioned above, which contrasts with the COSMOGLobe approach of assigning a linear baseline solution for the entire scan, and then accounting for the non-linearity through  $\mathbf{n}^{\text{corr}}$ . The variations are commensurate with the correlated noise correlation length, which for K113 has  $f_{\text{knee}} \sim 0.5$  mHz, corresponding to a



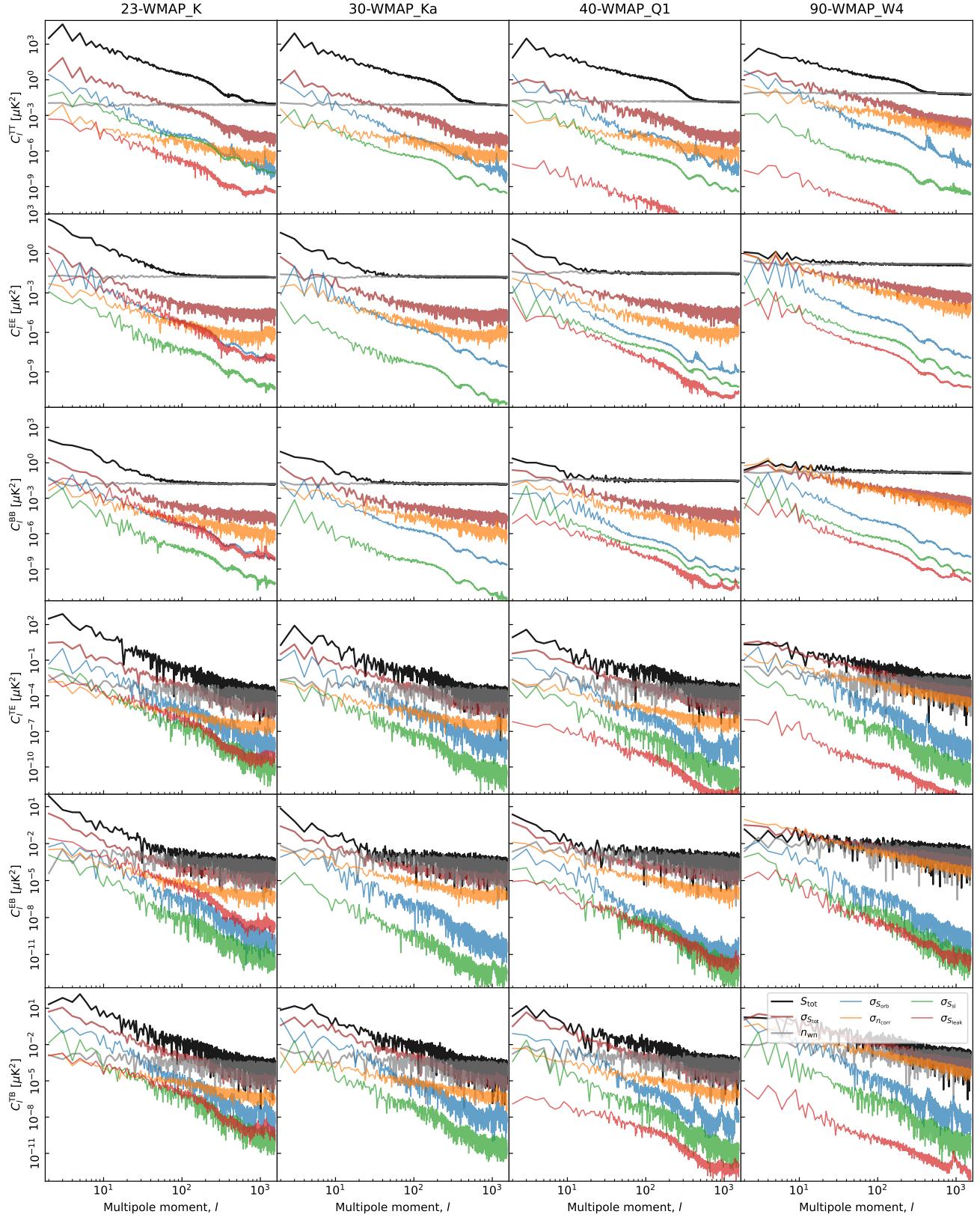
**Fig. 11.** TOD corrections for  $K$ -band for a single Gibbs sample, projected into maps. Columns show Stokes  $T$ ,  $Q$ , and  $U$  parameters. Rows show, from top to bottom, 1) correlated noise; 2) the orbital dipole; 3) bandpass mismatch leakage; and 4) sidelobe corrections. The bottom row shows the residual obtained when binning the sky and systematics-subtracted TOD into a sky map. Note that the correlated noise and residual have been smoothed by a  $2^{\circ}$  Gaussian beam.

little over half an hour. Therefore, the hourlong baseline subtraction essentially acts as a destriper, removing an estimate of the correlated noise.

#### 4.5. Instrumental corrections in map domain

Returning to the global parametric data model in Eq. (7), it is useful for intuition purposes to project each of the *WMAP* TOD-

level correction terms into sky maps. This is done for  $K$ -band in Fig. 11, in the same form as was done for LFI by Basyrov et al. (2022). Columns correspond to Stokes  $T$ ,  $Q$ , and  $U$  parameters, while rows show, from top to bottom, 1) correlated noise; 2) the orbital CMB dipole; 3) bandpass leakage; and 4) sidelobe corrections. The bottom row shows the residual obtained after subtracting all modelled terms from the raw TOD. All maps correspond to one single and randomly selected Gibbs sample.



**Fig. 12.** Pseudo-spectrum standard deviation for each instrumental systematic correction shown in Fig. 11 (thin colored lines). For comparison, thick black lines show spectra for the full coadded frequency map; thick red lines show the standard deviation of the same (i.e., the full systematic uncertainty); gray lines show white noise realizations; and black lines show the power spectra of the maps themselves. Columns show results for  $K$ ,  $Ka$ ,  $Q1$ , and  $W4$ , respectively, while rows show results for each of the six polarization states ( $TT$ ,  $EE$ ,  $BB$ ,  $TE$ ,  $TB$ , and  $EB$ ). All spectra have been derived outside the CMB confidence mask presented by Andersen et al. (2022) using the HEALPix anafast utility, correcting only for sky fraction and not for mask mode coupling.

Starting with the correlated noise in the top row and the residual in the bottom row, we note that these are the only terms that are fundamentally stochastic in nature; all the other terms are primarily deterministic in nature, as they rely only on the sky model coupled to a small number of instrumental parameters, such as the gain, bandpass, or beam. As such, the correlated noise and residual maps essentially act as the “trash cans” of Bayesian CMB analysis; together they show anything in the TOD that is not explicitly modelled. For the  $K$ -band channel, we see that the correlated noise is limited to less than  $1 \mu\text{K}$  over almost the full sky, while the residual appears like random noise over most of the sky. The main exceptions to this are strong residuals near the Galactic plane in temperature, which indicates the presence of unmodelled foreground features; typical candidates to explain this would be curvature in the synchrotron spectral index or a more complicated AME SED than that assumed here. Secondly, at high latitudes we see traces of a small dipole with an amplitude of  $1\text{--}2 \mu\text{K}$  aligned with the CMB dipole; and this indicates the presence of an absolute calibration deviation of about  $0.03\text{--}0.06\%$ ; this is within the uncertainty of the absolute  $K$ -band calibration prior of  $0.1\%$  discussed in Sect. 2.5, and when inspecting different Markov chain samples of the same type, one can see that the amplitude and sign of this residual scatter around zero as expected.

Next, the amplitude of the orbital dipole is about  $250 \mu\text{K}$  in temperature and  $2.5 \mu\text{K}$  in polarization. This component by itself is exceedingly well known, as it depends only on the satellite velocity and the CMB monopole temperature. However, when actually fitting this to the real data, it obviously also depends on both the gain, and sample-to-sample variations in this map is therefore a good tracer of gain uncertainties. In addition, its physical appearance also in principle depend on the sidelobe model, but we do not yet propagate sidelobe uncertainties anywhere in the analysis.

The third row shows the corrections for bandpass leakage. (In principle, this also accounts for leakage due to beam differences, as discussed by Svalheim et al. (2022a), but for WMAP this effect is much smaller than the bandpass effect.) As for *Planck* LFI, this term is by far the largest correction in polarization, with an amplitude that is almost an order of magnitude larger than any of the others. It is highly non-trivial to compare this component to the WMAP9 analysis, since this effect was accounted for by solving for a spurious  $S$  map as part of their mapmaking. However, it is important to note that whether one models this effect explicitly, as we do, or project it out during mapmaking, as the WMAP pipeline did, the accuracy of the bandpass correction depend directly on the accuracy of the gain and transmission imbalance calibration.

The fourth row shows the sidelobe contribution. Here we see that the temperature amplitude reaches  $50 \mu\text{K}$ , which corresponds to  $1.5\%$  of the CMB Solar dipole amplitude. If it should turn out that the sidelobe model were incorrect by, say,  $30\%$ , this translates directly into an error in the absolute  $K$ -band calibration of about  $0.5\%$ , which is significantly greater than the statistical error shown above. Given the degeneracies discussed in Sect. 2.5, there is thus also a strong degeneracy between the AME dipole and the  $K$ -band sidelobe. For polarization, the amplitude is mostly smaller than  $1 \mu\text{K}$ , and therefore of minor importance for this channel. (We note that only the intensity component of the WMAP sidelobe model has been published, and polarized sidelobes are currently not accounted for in the current processing. However, as shown by Barnes et al. (2003), the amplitude of this contribution is relatively small, and typically less than  $\lesssim 0.5 \mu\text{K}$  over most of the sky.)

#### 4.6. Instrumental uncertainties in power spectrum domain

We conclude this section by estimating the uncertainty of each instrumental effect in terms of angular power spectra, using the same methodology as Basyrov et al. (2022) for *Planck* LFI. That is, we first compute the power spectra for each individual instrumental correction Markov chain sample, as illustrated in Fig. 11, and then compute the standard deviation of all these samples. The results from this computation are summarized in Fig. 12 for four DAs, namely  $K$ ,  $Ka$ ,  $Q1$ , and  $W4$ . In each panel the black line shows the power spectrum of the full co-added frequency sky map (including both signal and noise), while the gray line shows the white noise level. The thick dark red line shows the sum of all variations, while the thin colored lines show the contribution from individual correction terms.

On large angular scales in the  $TT$  spectrum, we see that the dominant uncertainty comes from the orbital dipole (blue lines), which essentially trace gain uncertainties. This makes intuitively sense, since these data are strongly signal-dominated; indeed, for  $K$ -band even the sidelobe contribution (green lines) is higher than the correlated noise.

For large-scale  $EE$  polarization, the dominant term varies from channel to channel. Specifically, for  $K$ -band the bandpass leakage (thin red lines) and gain fluctuations are significantly larger than the correlated noise, while for  $Q1$  and  $W$ -band the correlated noise dominates for most multipoles, although the gain fluctuations are comparable for some  $\ell$ 's.

An important conclusion to be drawn from these measurements is that a simple uncertainty model that primarily accounts for correlated noise is likely to be sub-optimal for detailed cosmological analysis of large-scale polarization. Both gain and bandpass uncertainties are at least as important for the lowest multipoles, and simultaneously accounting for all of these contributions is clearly important in order to derive robust cosmological results.

## 5. Frequency maps

In this section, we present the reprocessed WMAP frequency maps and their properties. In Sect 5.1 we present the reprocessed WMAP maps themselves, commenting on notable features. Section 5.2 compares the properties of the individual DAs with the published WMAP9 results, while Sect. 5.3 focuses on the internal consistency between the WMAP channels themselves. Finally, we assess the consistency between the LFI channels and WMAP in Sect. 5.4 and compare with legacy results.

### 5.1. Map summary statistics

For each DA, we present the mean maps, the white noise contribution in the form of RMS maps, and the standard deviation of the maps accounting for the sampling of instrumental parameters. Each mean map retains the Solar dipole, which we will estimate separately in Sect. 6.1.1. Additionally, we present inverse-weighted mean maps of  $Q$ ,  $V$ , and  $W$  rather than displaying each individual DA.

The  $K$ -band,  $Ka$ -band,  $Q$ -band,  $V$ -band, and  $W$ -band mean maps are presented in Figs. 13, 14, 15, 16, and 17, respectively, given in  $\mu\text{K}_{\text{CMB}}$  units. The  $Q$ ,  $V$ , and  $W$ -band maps are generated by inverse-variance weighting the individual DAs. The temperature maps are presented at full resolution, while the polarization maps have been smoothed with a  $2^\circ$  Gaussian beam. The maps in temperature behave as expected, with consistent Solar

dipole and Galactic foreground emission decreasing as the frequency increases. Similarly, the polarized maps decrease from  $K$ - $V$ -band following the expected synchrotron behavior, with a slight increase at  $W$ -band due to the contribution of thermal dust. Most striking, especially when compared with the delivered *WMAP9* maps, is the lack of transmission imbalance modes or poorly measured modes in the polarized maps.

The RMS maps are computed during the mapmaking routine, adding  $\sigma_{0,i}^{-2}$  for every timestep that horn is in a given pixel, appropriately scaled by polarization angle and imbalance parameters. The top row of Fig. 18 shows the white noise for the Stokes parameters and the correlation coefficient between Stokes  $Q$  and  $U$ . The white noise pattern for  $T$  follows the usual pattern with highest sensitivity at the North and South ecliptic poles, as well as circles around the poles corresponding to times when the partner horn is observing the opposite ecliptic pole. There are also regions of higher noise level corresponding to planets crossing the ecliptic, and regions of higher emission  $\approx 140^\circ$  away from the Galactic center, which correspond to the times when the partner horn lies within the processing mask.

The polarized RMS maps share all of these characteristics, but with an overall amplitude shift due to polarization measurements having half the effective number of observations per pixel. In addition, the poles have a characteristic “X”-like structure that is rotated  $45^\circ$  degrees between  $Q$  and  $U$ , corresponding to different polarization orientations. There are also characteristic large scale structures visible in Galactic coordinates, corresponding to polarization modes poorly constrained by the *WMAP* scan strategy.

While the maps in the top row of Fig. 18 are directly comparable to the *WMAP9* products, the bottom row shows a unique product, the standard deviation and  $QU$  correlation of the output chain maps. These maps can be considered the “systematic” error contributions, as their variation depends on the sampled instrumental parameters, i.e., gain, imbalance parameters, correlated noise, and sidelobe correction. The temperature map contains a clear quadrupole signature. This is due to the variation in the absolute calibration  $g_0$ , which changes the Solar dipole in the final map. In addition to the quadrupole, the Galactic plane also varies due to the gain solution being varied. As expected, the white noise patterns associated with the scan strategy also appear in the polarization maps, and become more clear in the temperature maps for the higher frequency DAs (Fig. B.3). As these maps only include 490 total samples out of full 500-sample Gibbs chain, they will be superseded by a larger future chain. While the quantitative solution will change, it is unlikely that the final maps will differ qualitatively from those presented here.

An additional useful term to consider is the difference between two arbitrary samples, which we show in Fig. 19. In temperature, the most clear term is a dipole corresponding to the absolute gain difference and the Galactic plane. There are also additional small lines associated with the scanning strategy, which correspond to different correlated noise realizations. In polarization, correlated noise is the dominant difference between two samples, with a small imprint of the Galactic plane due to relative gain variation. The polarization differences are aligned with *WMAP*’s scans, modulated by the polarization angle.

We also consider the spatial structure of the TOD corrections in pixel space, as shown in Fig. 11. This corresponds to the TOD objects in Fig. 1 binned into Stokes  $I$ ,  $Q$ , and  $U$  maps. Note that the dynamic range of each of these figures varies by two orders of magnitude, indicating that some components must be subtracted more carefully than others. However, the amplitude of the signal is not proportional to the level of scrutiny necessary,

as some, such as orbital dipole, are much more precisely known than the others.

The variation of these TOD corrections is more important than the absolute amplitude. As shown in Fig. 12, white noise is the dominant effect for  $\ell \gtrsim 100$  for all frequencies. For temperature, the low-multipole variation in the orbital dipole signal and far sidelobe corrections are the largest source of fluctuations and low multipoles, but remain orders of magnitude below the average signal. These fluctuations are directly related to absolute calibration uncertainty, and provide a practical limit to the magnitude gain effects the temperature signal. For the  $C_\ell^{\text{EE}}$  and  $C_\ell^{\text{BB}}$  spectra, the relative amplitude of the fluctuations are band-dependent. For example, bandpass leakage and orbital dipole corrections are the dominant effect, while for  $Q$ -band, the correlated noise realizations contribute the most variation. In general, we find that for large angular scale polarization, modelling uncertainties in the form of gain and sidelobe estimation have a comparable effect to white noise and correlated noise estimates.

As a final inspection of the COSMOGLOBE map products, we take the angular power spectra of the maps themselves and compare with the *WMAP9* results in Fig. 20. To compute the power spectrum, we mask the data using the extended temperature analysis mask which allows a sky fraction of 68.8 % and obtain the pseudo- $C_\ell$  power spectra using the NaMaster (Alonso et al. 2019)<sup>7</sup> `compute_full_master` routine, returning a set of decoupled bandpowers.

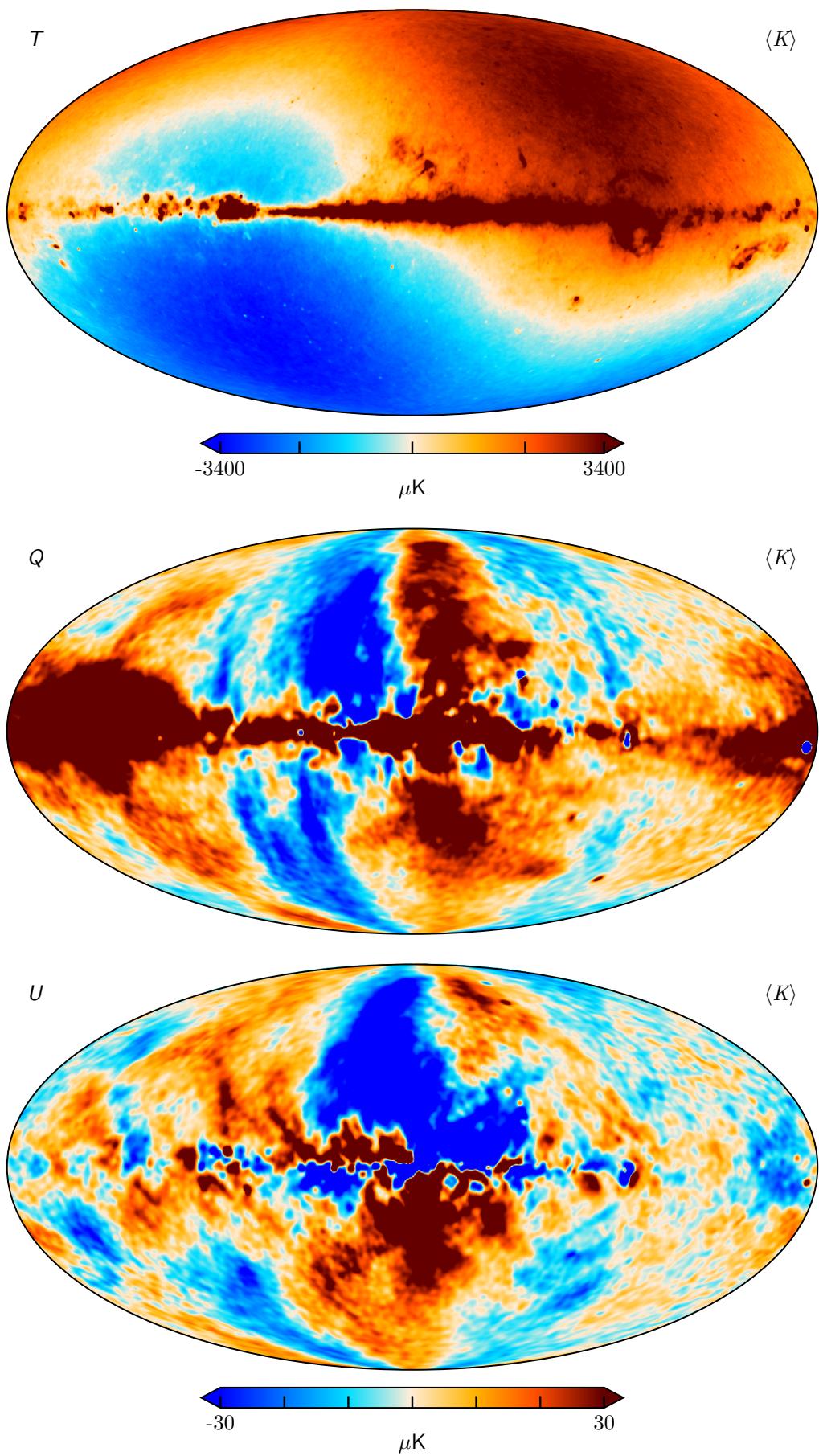
As the temperature power spectra are signal-dominated up to  $\ell \sim 200$  for all DAs, it is more useful to look at the ratio of spectra in the left column of Fig. 20. Here we see that the spectra are consistent with each other at all but the very largest and smallest scales. The largest scale differences are due to mode-coupling effects – the *WMAP9* maps were produced by removing the Solar dipole in the timestream before mapmaking, while the COSMOGLOBE maps needed a dipole estimate to be removed as a post-processing step. The small scale differences above  $\ell \sim 200$  can be attributed to the white noise treatment. In particular, the *WMAP9* processing gain solution varies every 23 s compared to the COSMOGLOBE solution which uses constant gain per scan. Conversely, the COSMOGLOBE radiometer noise estimate varies per scan, whereas there is no mention of raw instrumental noise variation with time in the *WMAP* suite of papers.

The  $E$ -mode power spectra, displayed in the second column of Fig. 20, are mainly dominated by noise and polarized synchrotron emission. As expected, the large scale foreground-dominated multipoles decrease in amplitude according to the relative amplitude of the synchrotron spectrum. Aside from the  $\ell = 8$   $W2$  multipole, the COSMOGLOBE power spectra are well-behaved across all bands. The large fluctuations in the *WMAP* power spectrum, in particular  $W2$  and  $W4$ , are almost completely gone in the COSMOGLOBE analysis.

The  $B$ -mode power spectra, displayed in the third column of Fig. 20, should follow the same pattern as in the  $E$ -modes, but with foregrounds reduced by a factor of  $\approx 2$  (Bennett et al. 2013). Indeed, with the notable exception of the  $Ka$  and  $Q1$   $\ell = 3$  multipoles and the  $W3$   $\ell = 7$  multipole, this pattern is largely borne out, with nearly white noise across all angular scales. The  $C_{\ell=3}^{\text{BB}}$  mode has been identified as being poorly measured by, e.g., Jarosik et al. (2011), due to its symmetry aligning with  $\gtrsim 10$  min signals in the TOD induced by the *WMAP* scan strategy.

In general, we find that the reprocessed *WMAP* maps produced in the COSMOGLOBE framework have map and power spectrum properties that contain traces of the *WMAP* observation

<sup>7</sup> <https://github.com/LSSTDESC/NaMaster>



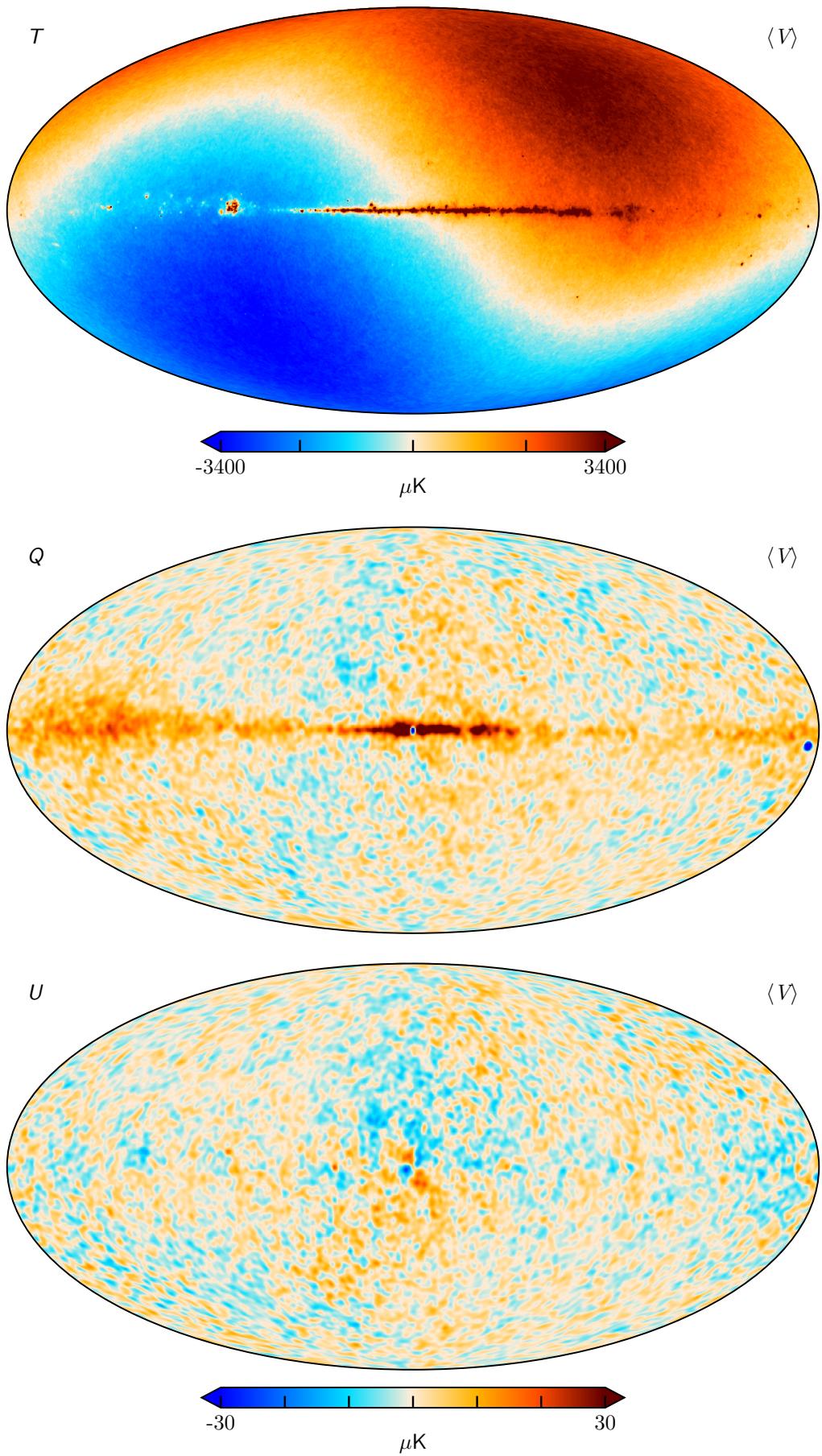
**Fig. 13.**  $K$ -band



**Fig. 14.**  $Ka$ -band



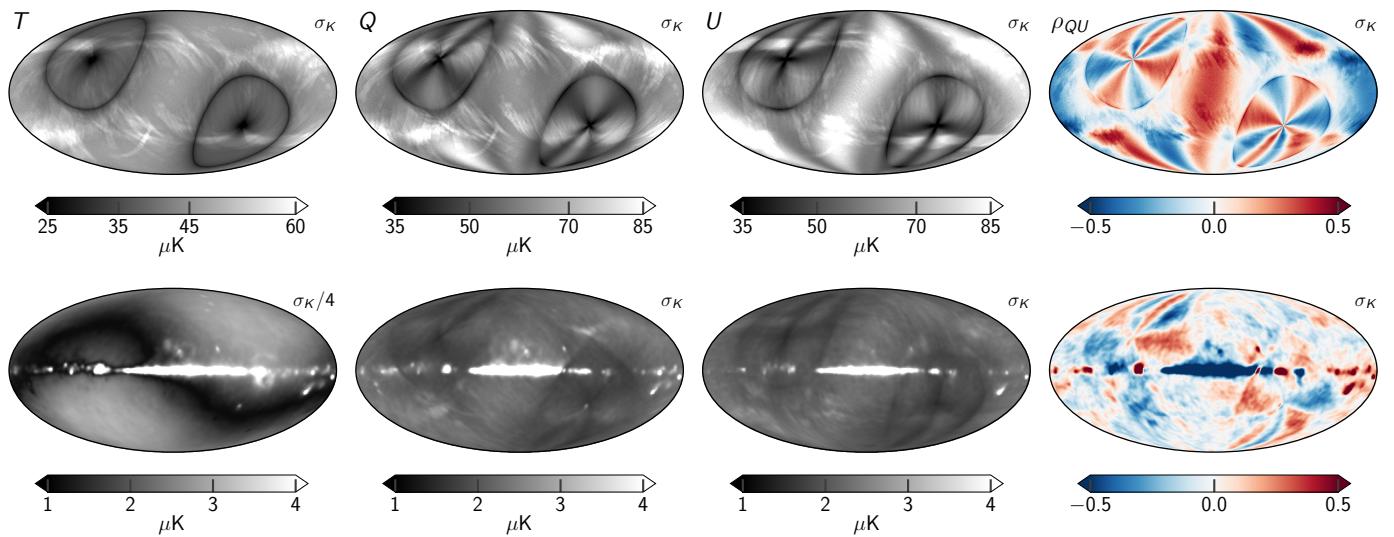
**Fig. 15.**  $Q$ -band



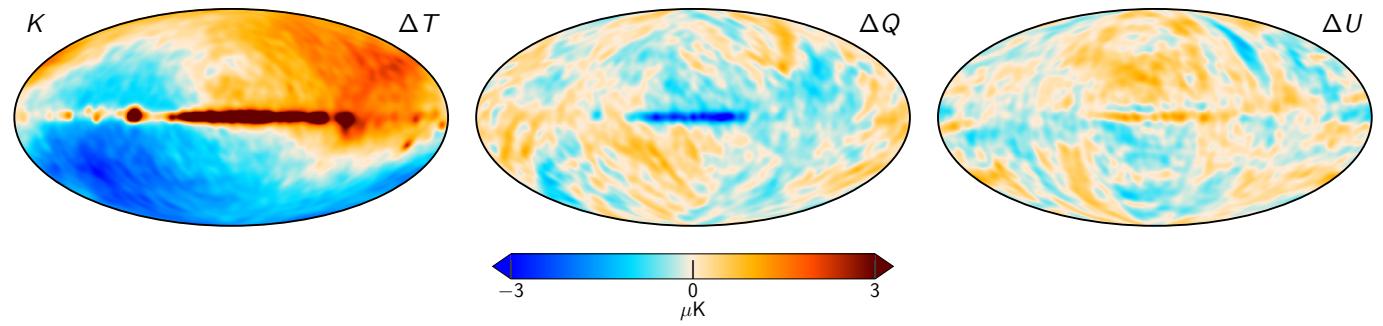
**Fig. 16.**  $V$ -band



**Fig. 17.**  $W$ -band



**Fig. 18.** Posterior variation maps for  $K$ -band. Columns show the Stokes parameters and the correlation coefficient between  $Q$  and  $U$ , while the rows show (top) the white noise rms per pixel and (bottom) the posterior standard deviation. The rms maps are unsmoothed, while the standard deviations have been smoothed to  $7^\circ$ .



**Fig. 19.** Difference between two  $K$ -band Gibbs samples, smoothed to  $7^\circ$ .

strategy than the *WMAP9* products. To fully assess the quality of these frequency maps compared to the official *WMAP9* products, we compare the maps explicitly in Sect. 5.2.

## 5.2. Comparison with 9-year WMAP maps

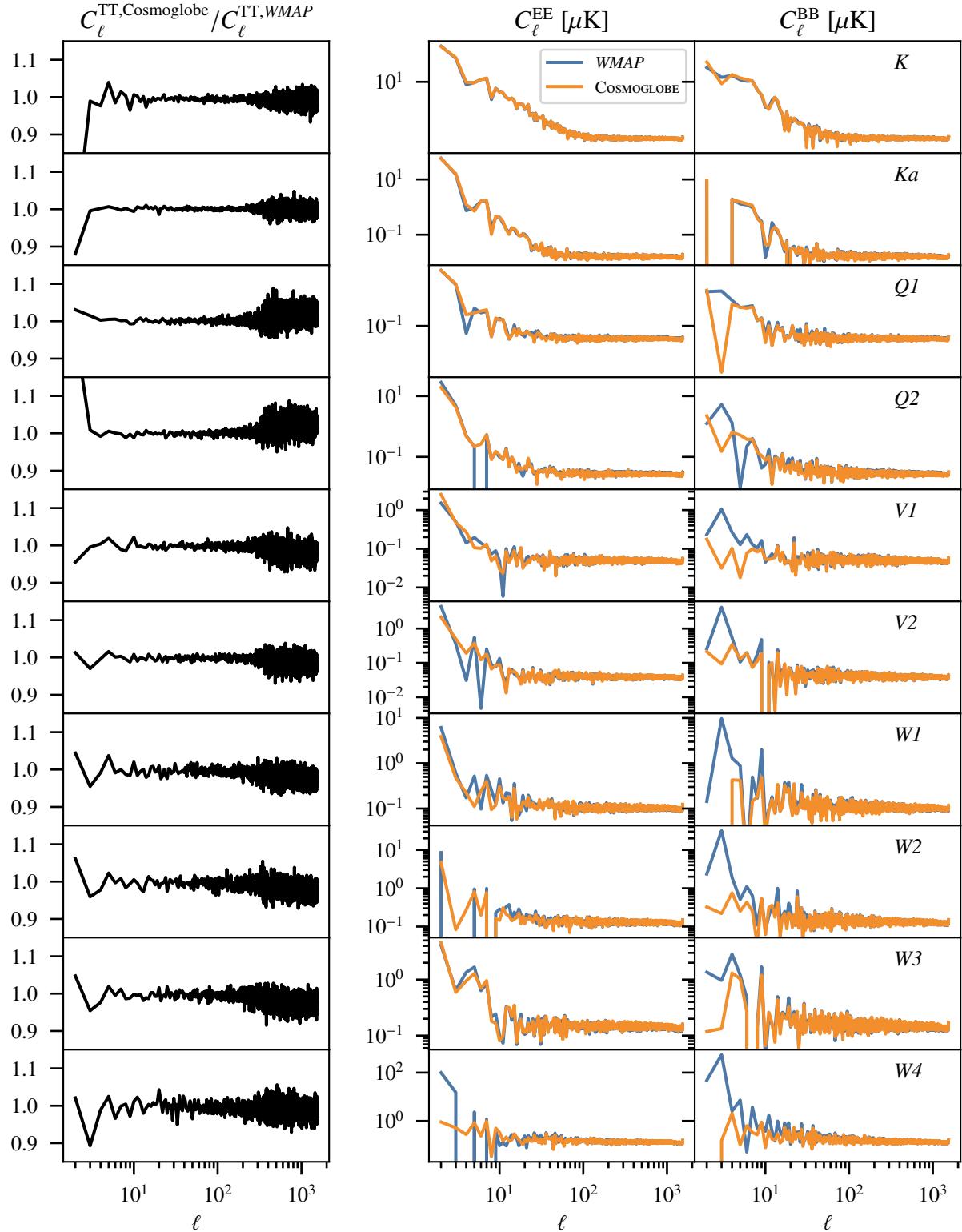
We present difference maps between the official 9-year *WMAP* maps and the maps produced in this work in Fig. 21. This figure shows a total of fifteen difference maps, one for each of the main *WMAP* bands ( $K$ ,  $Ka$ ,  $Q$ ,  $V$ , and  $W$ ), in the three Stokes parameters.

The differences are overall quite small, as the data processing between *COSMOGLOBE* and *WMAP9* are quite similar, with subtle differences as described in Sec. 3. In total intensity, we see good agreement with the full *WMAP9* results, with deviations at the few  $\mu\text{K}$  level. The largest difference between the two analyses is shown in the  $K$ -band difference map, which demonstrates the difference between the band calibration as described in Sects. 4.2 and 7.3. This can be attributed to the absolute calibration differences between the two pipelines. In particular, Bennett et al. (2013) estimate a calibration uncertainty of 0.2 % across all bands, corresponding to a  $\sim 7 \mu\text{K}$  variation in the Solar dipole amplitude. On the other hand, the *COSMOGLOBE* absolute calibration prior width of 0.002 can induce a  $6 \mu\text{K}$  Solar dipole residual. Therefore, a dipole difference such as this is not unexpected. In bands  $Ka-W$ , the absolute calibration is driven wholly

by the sampling algorithms described in Sect. 2.7.1, and as such is much more susceptible to poor modeling of the data. The fact that the dipole amplitude in the difference maps is so small is strong evidence that the parametric modeling of the absolute calibration performance is comparable to the *WMAP9* approach of modelling the gain using on-board thermistors.

In the  $V$  and  $W$ -band temperature maps, there is an additional quadrupole signal closely aligned with the Solar dipole. As noted by Larson et al. (2015), the *WMAP9* maps retain the kinematic quadrupole, whereas *Commander3* removes this component. While kinematic quadrupole has the expected shape, the frequency dependence is not consistent with the expected functional form  $x \coth x$  where  $x = h\nu/(2kT_{\text{CMB}})$  (Notari & Quartin 2015). It is therefore likely that the quadrupole difference comes from some second-order effects in the time variation of the gain. There is indeed an oscillatory structure in the relative difference between the gain solutions, which can be most seen in Fig. A.4.

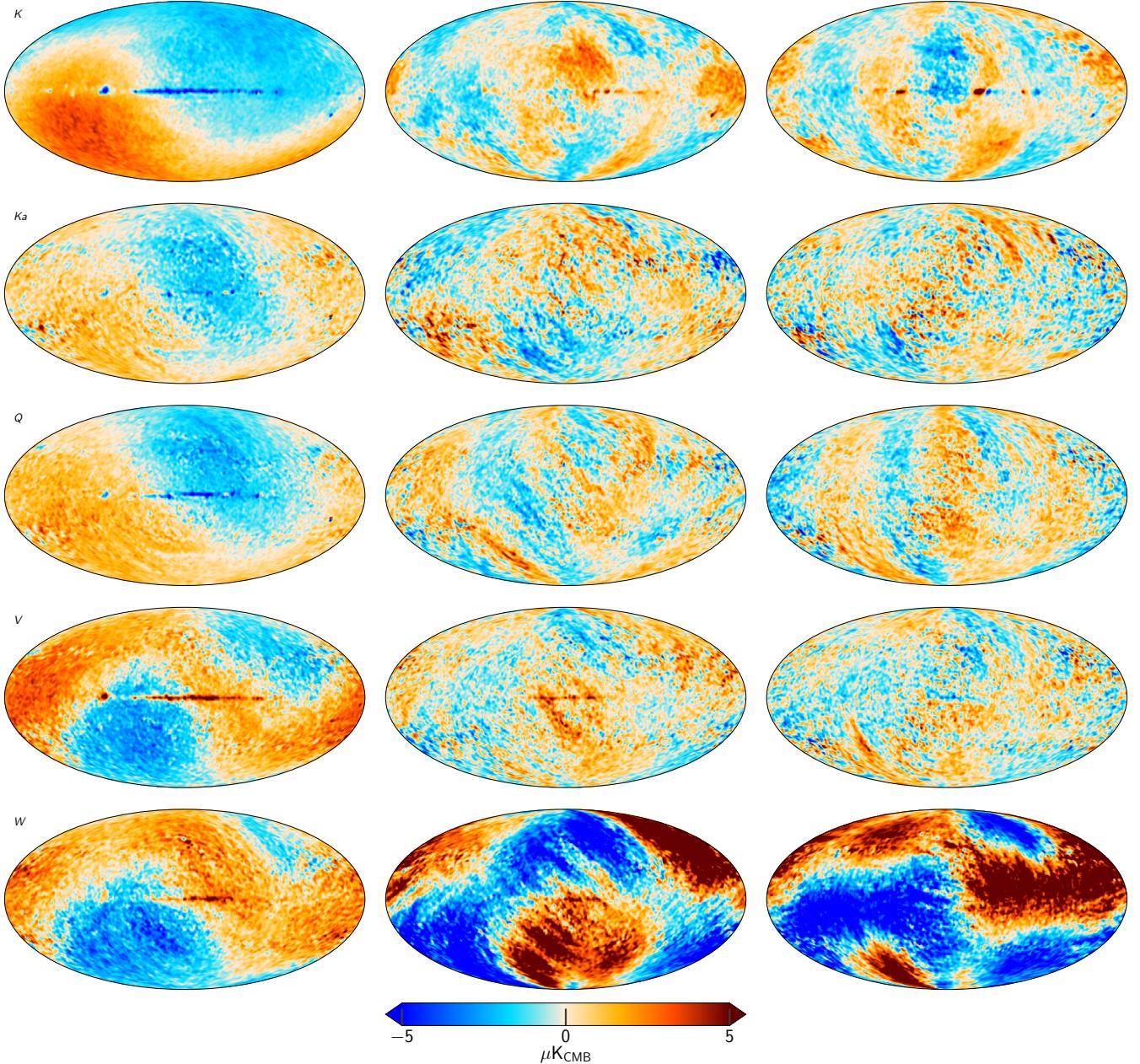
In polarization, we note large scale differences in both Stokes  $Q$  and  $U$ . These large scale differences do not match known Galactic component morphologies. All of these differences are well-described by linear combinations of the poorly measured modes, though the map-space morphologies are not identical. These large mode differences are due to three main effects: 1) poor polarization angle coverage for a few large-scale modes; 2) errors in transmission imbalance coupled with the Solar dipole; and 3) interplay between the transmission imbalance, the far



**Fig. 20.** Comparison of the  $C_{\ell}^{TT}$ ,  $C_{\ell}^{EE}$ , and  $C_{\ell}^{BB}$  from WMAP9 and COSMOGLOBE. Each row corresponds to a different DA, with frequency increasing from top to bottom. (left): ratio of  $C_{\ell}^{TT}$  from COSMOGLOBE compared to WMAP9. (middle/right):  $C_{\ell}^{EE/BB}$  power spectra with WMAP9 in blue and COSMOGLOBE in orange.

sidelobe, and the Solar dipole, as briefly described in Sect. ???. The scale of these effects is most pronounced in the W-band polarization results, where we see the largest differences between the two processing pipelines.

The removal of the poorly measured modes demonstrates the benefit of jointly processing these two experiments simultaneously in the same framework. The differences shown in Fig. 21, particularly in polarization, are due to effects in the WMAP9 maps that cannot be constrained by the WMAP observation strat-



**Fig. 21.** Difference maps between the COSMOGLobe and 9-year WMAP frequency maps. Columns show Stokes  $T$ ,  $Q$ , and  $U$  parameter maps, while rows show  $K$ -,  $Ka$ -,  $Q$ -,  $V$ -, and  $W$ -band maps. The maps are all smoothed with a  $2^\circ$  FWHM Gaussian beam.

**Table 5.** Difference map  $\chi^2$  statistics.

DIFFERENCE	$\chi^2_{\text{uncorr}}$	$\chi^2_{\text{corr}}$	$\Delta\chi^2$
$0.32 \times K1 - Ka1 \dots$	4291	4287	4
$Q1 - Q2 \dots \dots \dots$	4500	4380	120
$V1 - V2 \dots \dots \dots$	4490	4429	61
$W1 - W2 \dots \dots \dots$	4328	4270	68
$W3 - W4 \dots \dots \dots$	4257	4145	112

egy. Instead, in the joint analysis, the full sky model allows for a framework in which large scale modes, clearly not Galactic in origin, to be attributed to the combination of imbalance parameter and Solar dipole uncertainty.

### 5.3. Consistency within WMAP channels

An important test for how well the instrumental systematics are being modeled is by checking the agreement within each of the WMAP channels. As described in Sect. ??, the  $Q$ -band and  $V$ -band channels each had two DAs, while the  $W$ -band had four DAs. Checking for discrepancies between each of the individual DA maps within the same frequency channel can highlight mismodeled systematics. While the  $K$ -band and  $Ka$ -band have different central frequencies, they are close enough that we can compare them by scaling  $K$ -band assuming a polarized synchrotron power law of  $\beta_s = -3.1$ .

The interchannel maps for Stokes  $Q$  and  $U$  are shown in Fig. 22. For the  $K-Ka$  difference, there are coherent signals in both the WMAP9 and Cosmoglobe difference maps. An unsurprising difference is along the Galactic plane, due to the sky not

**Table 6.** Transmission imbalance template amplitudes for each *WMAP* radiometer as estimated by fitting the official templates to low-resolution difference maps between COSMOGLOBE and *WMAP*. The templates are provided in mK, and the template amplitudes are therefore dimensionless. The fourth column lists the relative decrease in standard deviation,  $\sqrt{\sigma_{\text{raw}}^2 - \sigma_{\text{corr}}^2}/\sigma_{\text{raw}}$ , after subtracting the best-fit templates in percent.

DA	$a_1$	$a_2$	$\Delta\sigma[\%]$
K1 .....	-27.5	-50.6	30
Ka1 .....	-1.4	-1.9	25
Q1 .....	-30.0	-71.6	11
Q2 .....	-7.1	-1.5	20
V1 .....	-32.8	-53.4	6
V2 .....	8.8	-4.1	16
W1 .....	-2.8	4.6	8
W2 .....	-6.9	-3.5	11
W3 .....	29.1	53.4	12
W4 .....	15.5	-6.8	52

being well-described by a simple global power law in this region. However, the large-scale Stokes  $Q$  map has a large scale residual that is not matched in  $U$ . This coherent structure is present in both the *WMAP9* and COSMOGLOBE maps, although in *WMAP9* the structure is interrupted by the poorly measured modes. In comparison, the COSMOGLOBE  $U$  map has much less large-scale structure than the *WMAP9* difference map.

The internal DA differences provide a cleaner comparison, with no astrophysical assumptions required. The  $Q$ -band and  $V$ -band half-difference maps from COSMOGLOBE have virtually no trace of the poorly measured modes, and the difference patterns are well-traced by the rms maps, with smaller differences at the ecliptic poles and larger differences along the ecliptic plane. The largest visual improvement is in the  $W$ -band half-difference, where the COSMOGLOBE case is almost entirely consistent with white noise, as opposed to the *WMAP9* difference that is dominated by the poorly measured modes. Based on these visual comparisons, the COSMOGLOBE maps themselves contain essentially none of the poorly-measured modes and are more self-consistent than the *WMAP9* frequency maps.

As noted by Jarosik et al. (2011), the low- $\ell$   $W$ -band polarization data were excluded from cosmological analysis due to excess variance in the  $\ell \leq 7$  multipoles. To test the COSMOGLOBE maps' performance at these scales, we take the power spectrum of the full-sky difference maps using the standard anafast routine in Fig. 23. With very few exceptions, the *WMAP9* power spectra have much more power at  $\ell \leq 7$  than the COSMOGLOBE maps in both the  $E$ -modes and  $B$ -modes. Of particular note is the  $\ell = 3$   $B$ -mode, which has consistently been identified as poorly measured in the *WMAP* scan strategy, and has been reduced in every difference spectrum. Based on these power spectra, we can identify no reason to exclude the reprocessed  $W$ -band polarization data in future cosmological analyses.

#### 5.4. Consistency between WMAP and LFI

As *WMAP* and LFI observe an overlapping frequency range, it is vital to assess how well these two experiments agree with each other, as the sky they observe is the same. Inspecting comparisons of the two experiments' sky maps can help elucidate differences between them induced by scan strategy and instrument design. Figure 24 shows comparisons between the *WMAP*  $K$ - and  $Ka$ -bands and the LFI 30 GHz channel, between the *WMAP*

$Q$ -band and LFI 44 GHz, and finally between *WMAP*  $V$ -band and LFI 70 GHz. For demonstration, the sky maps produced in this work are compared against the official *WMAP9* and BEYOND-PLANCK maps. Additionally, we compare the mean  $W$ -band maps with the DR4 100 GHz channel. In this comparison, the HFI map has had no input from Commander3 so this difference map is an independent comparison between two datasets and processing methods.

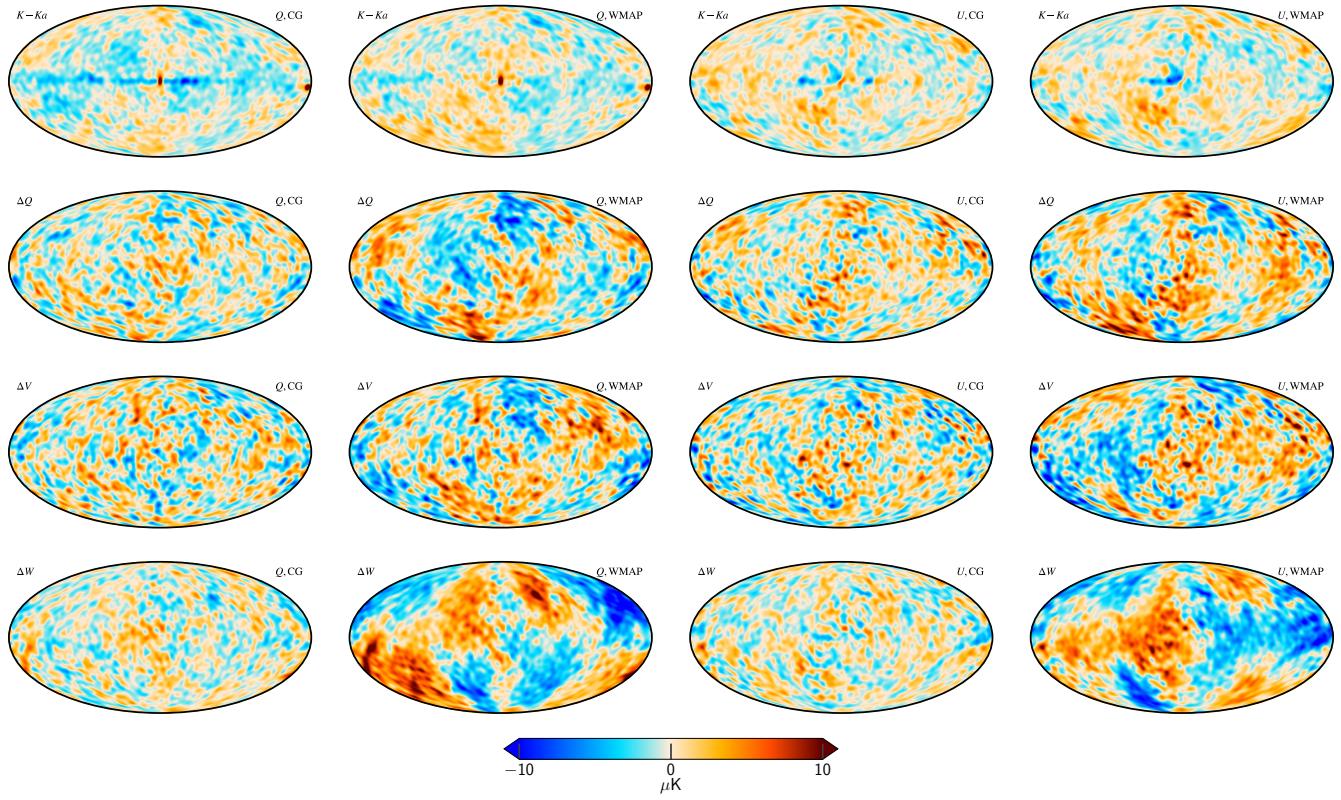
Starting with the COSMOGLOBE maps, we see in the first and third columns of Fig. 24 that the magnitude of the differences are small in both Stokes  $Q$  and  $U$ . Overall, across all five frequency map comparisons we see small levels of variation, with structure contained to the Galactic plane. Notably, however, there is a larger sky signal within the  $Ka$  – 30 Stokes  $Q$  comparison. This deviation is consistent with that from the  $Ka$  –  $K$  comparison in Fig. 22. The fact that this difference persists between these two difference maps indicates that this large-scale signal is confined to  $Ka$ , and is due either to data processing artifacts unique to this DA or to unmodeled sky polarization in the  $Ka$  bandpass. This large-scale difference also exists in the  $Q$  – 44 Stokes  $Q$ , but did not appear in the internal  $Q$  half-difference map. This strengthens the hypothesis that the signal is on the sky, but could still be a result of *WMAP* data processing, as the signal consistently appears in the *WMAP* data. Overall, across all frequency map comparisons, deviations at high Galactic latitudes are low, demonstrating a robust ability to remove poorly measured modes from both experiments thanks to the joint processing. The clearest improvement in reprocessing is shown in the final row of Fig. 24. The structure of the COSMOGLOBE difference maps in both Stokes  $Q$  and  $U$  shows very low levels of variation, in both the Galactic plane and at high latitudes.

Columns two and four of Fig. 24 show the differences between the official *WMAP9* and BEYONDPLANCK LFI frequency maps. Similar to the COSMOGLOBE sky map comparisons, we see differences in the Galactic center, and to a lesser degree along the Galactic plane due to the slight differences in the frequency coverage. When comparing the official *WMAP* maps, particularly for  $K$ -band, we see structures sweeping across large angular scales across the sky, likely due to the poorly measured modes in  $K$ -band. These structures were noted within the BEYONDPLANCK project, particularly in the polarized component separation and large-scale polarized AME studies (Svalheim et al. 2022b; Herman et al. 2022).

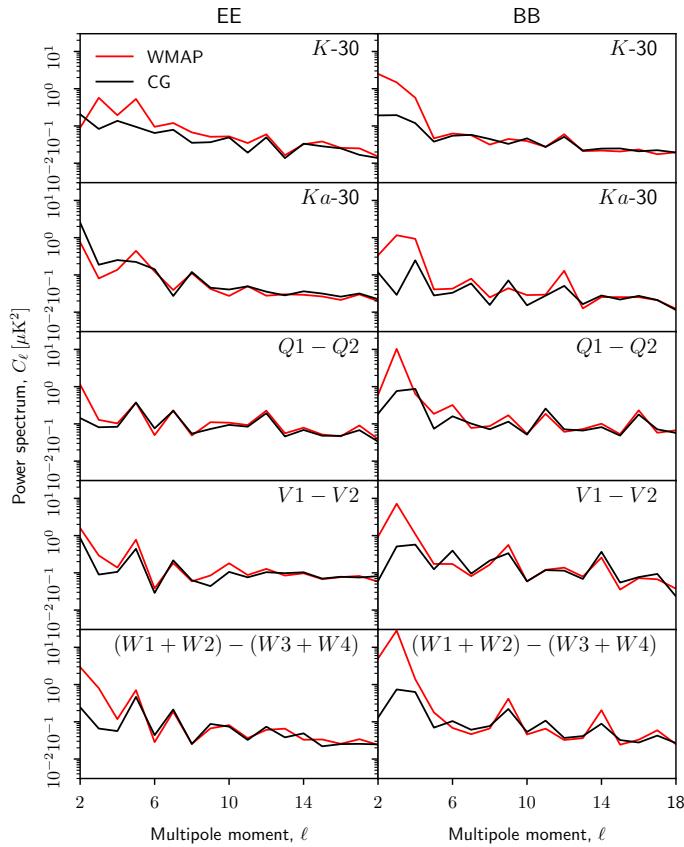
Of particular note is the 100 –  $W$  difference map. The COSMOGLOBE difference maps here have a similar level of white noise and Galactic contamination as the  $V$  – 70 maps, whereas the *WMAP9* differences are driven by the transmission imbalance modes, each with an opposite sign and magnitude. The difference between 100 GHz and  $W$  demonstrates that the good agreement between the *WMAP* and *Planck* LFI is not simply due to fitting low-level parameters in a joint analysis framework – by obtaining  $W$ -band maps that are consistent with an independent 100 GHz polarization map, we have shown that the *WMAP* is not simply the result of adding more free parameters to the fit, but a genuine improvement in data processing.

## 6. Preliminary astrophysical results

In this section, we present initial results for the astrophysical component separation temperature power spectra. The frequency coverage in this analysis is essentially the same as BEYOND-PLANCK, with the notable addition of the high signal-to-noise  $K$ -band and the reprocessed  $W$ -band. As such, the results presented



**Fig. 22.** Internal WMAP difference maps, smoothed by  $10^\circ$ . The two left columns are Stokes  $Q$ , and the two right columns are Stokes  $U$ , with the COSMOGLOBE and WMAP9 maps alternating between columns. The top to bottom rows are difference maps in increasing frequency.



**Fig. 23.** Full sky half-difference spectra. The red lines are the power spectra of the WMAP9 difference maps, while the black lines are the same for the reprocessed COSMOGLOBE maps.

here are similar in quality to the results presented by [Beyond-Planck \(2022\)](#).

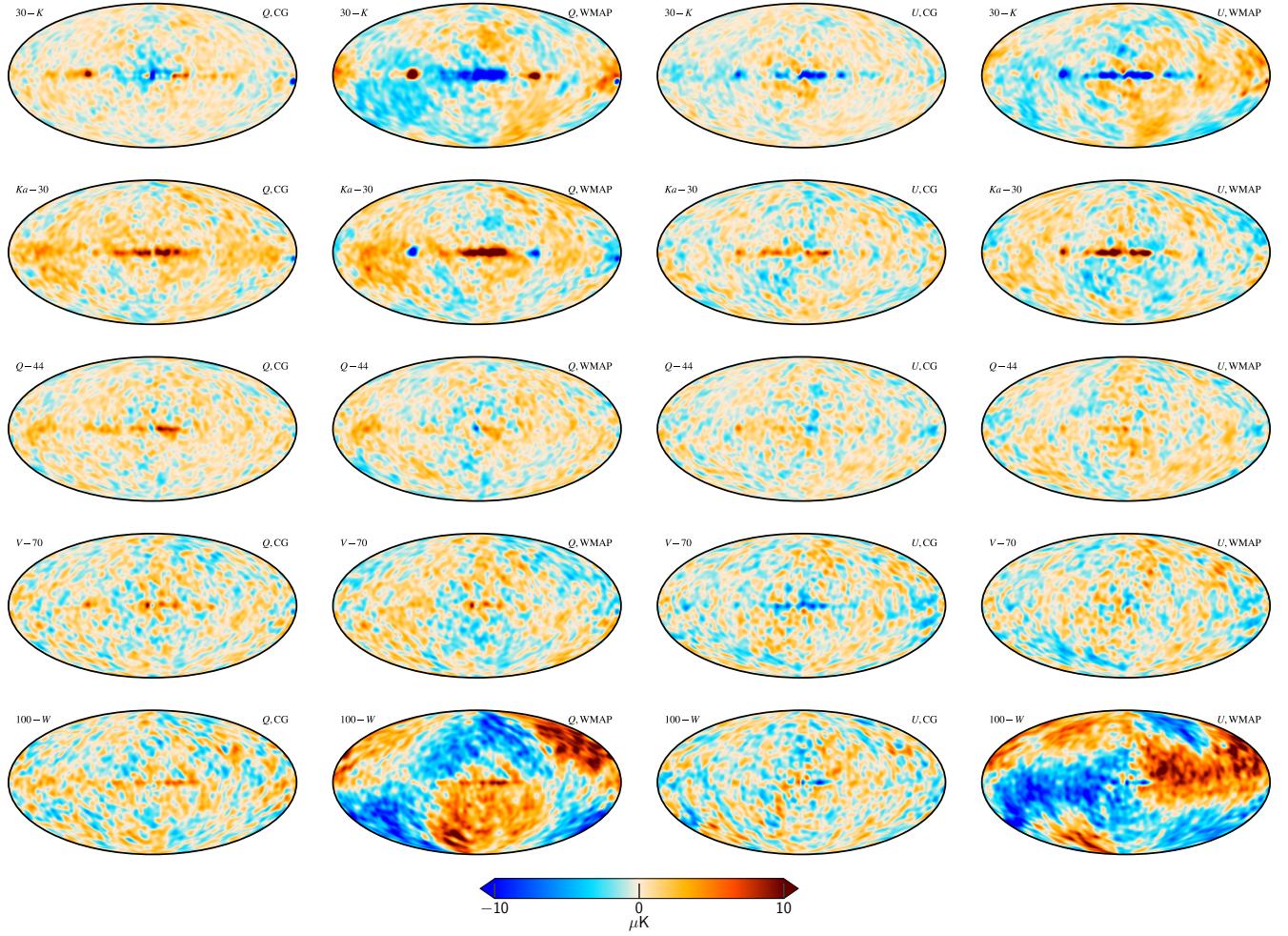
### 6.1. CMB results

Cosmological parameter estimation is left for future work, in large part because the two chains of length 250 are too short to reliably estimate cosmological parameters in this framework. For comparison, [Paradiso et al. \(2022\)](#) demonstrated that at least 2000 Gibbs samples were required before the reionization optical depth  $\tau$  value had converged. Similarly, for BEYONDPLANCK in the temperature case, the Gelman-Rubin statistic is just above  $R = 1.01$  for  $\ell \lesssim 600$  then continues to increase, indicating marginally acceptable convergence across all multipoles. Therefore, the results for the CMB presented here serve mainly as consistency checks. Given these caveats, we present preliminary CMB analysis in Sects. 6.1.1–6.1.2.

#### 6.1.1. Solar dipole

As argued by [Thommesen et al. \(2020\)](#), estimating the Solar dipole is one of the more difficult parameters to accurately constrain. This is in part due to the effect of mode-coupling when masking the Galactic plane, but also due to the Solar dipole in the calibration step. In essence, the Solar dipole's amplitude couples strongly to the calibration parameters. Misestimation of the calibration can propagate to an incorrect CMB dipole, and vice-versa.

The results of our Solar dipole estimate are displayed in Table 7 and Fig. 27. We find that the dipole direction is consistent with [Colombo et al. \(2022\)](#)'s result, but is  $11\,\mu\text{K}$  higher than the result from [Hinshaw et al. \(2009\)](#). Assuming a  $3400\,\mu\text{K}$  Solar



**Fig. 24.** Difference maps between similar WMAP and *Planck* frequency maps. The comparison plots go, by column: Stokes  $Q$  for the COSMOGLOBE sky maps, Stokes  $Q$  for official WMAP and BEYONDPLANCK data products, Stokes  $U$  for the COSMOGLOBEsky maps, and Stokes  $U$  for the official data products. (Top row) WMAP LFI 30 GHz minus  $K$ -band, scaled by the synchrotron power-law. (Top middle row) WMAP  $Ka$ -band minus LFI 30 GHz, also scaled by the synchrotron power-law. (Middle row) WMAP  $Q$ -band compared to the LFI 44 GHz sky maps, scaled by the synchrotron power-law. (Bottom middle row) WMAP  $V$ -band minus LFI 70 GHz, with unit scalings for each band. (Bottom row) The *Planck* DR4 100 GHz map minus the WMAP  $W$ -band also with unit scalings for each band.

dipole amplitude, an absolute calibration error of 0.3 % is sufficient to induce an  $11 \mu\text{K}$  error. In fact, the 0.2 % absolute calibration error reported by [Bennett et al. \(2013\)](#) can induce a  $6.7 \mu\text{K}$  Solar dipole amplitude error, dominating the error budget for the final reported Solar dipole.

It is also noteworthy that the BEYONDPLANCK Solar dipole differs from the COSMOGLOBE Solar dipole with a higher significance even than the WMAP9 result, despite using a nearly identical framework and almost identical datasets. First, we note that the *Planck* PR4 analysis ([Planck Collaboration LVII 2020](#)) showed a similarly apparently paradoxical discrepancy. On its face, the LFI 2018 and HFI 2018 Solar dipole values would average according to their uncertainties to  $\sim 3362.3 \mu\text{K}$ . In this analysis, [Planck Collaboration LVII \(2020\)](#) identify relative calibration uncertainty between 100 and 143 GHz as the dominant source of uncertainty.

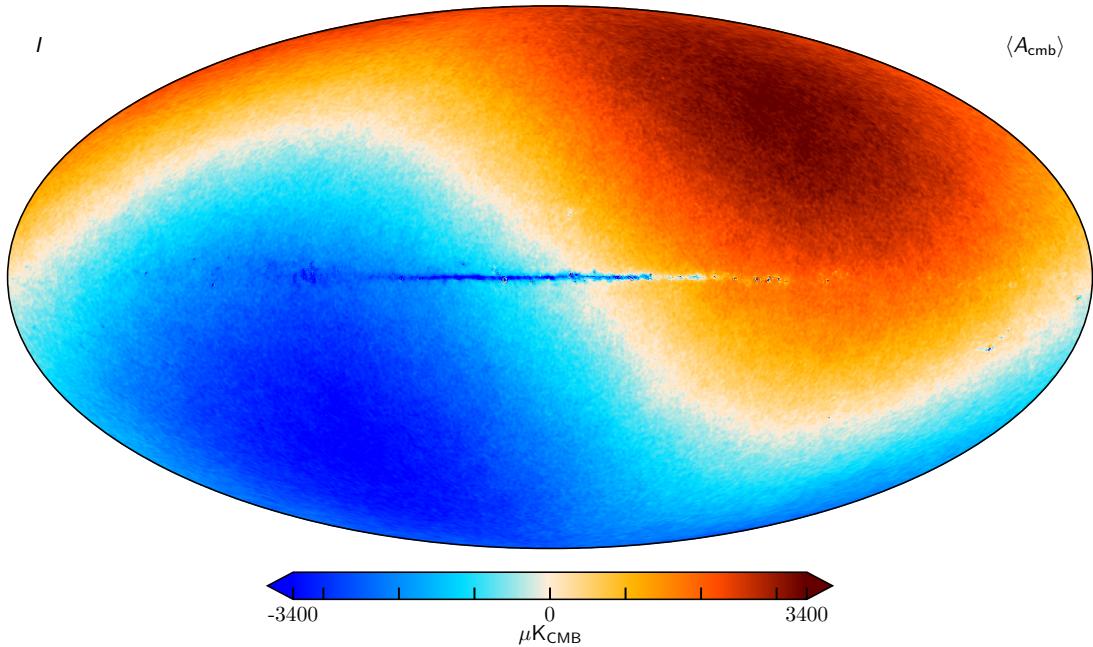
In the COSMOGLOBE analysis, we find that  $K$ -band has the highest signal-to-noise among all low frequency components, while being absolutely calibrated using a prior distribution. Therefore, it is likely that the chosen prior mean directly impacts the final Solar dipole amplitude, as it sets the amplitude of all low frequency components.

#### 6.1.2. Angular temperature power spectrum

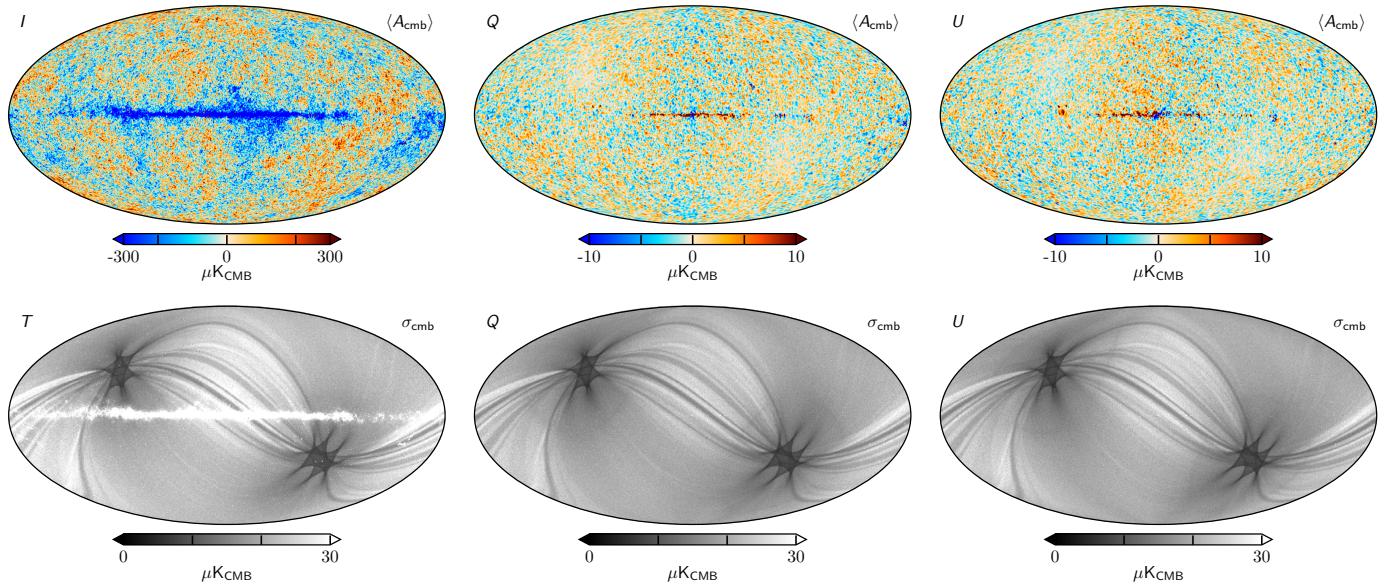
Figure 28 shows the angular temperature power spectrum derived from the CMB samples from the main Gibbs chain, obtained using a Blackwell-Rao estimator ([Chu et al. 2005](#); [Colombo et al. 2022](#)). We compare with the official WMAP ([Hinshaw et al. 2013](#)) and *Planck* ([Planck Collaboration V 2020](#)) power spectra, as well as the BEYONDPLANCK ([Colombo et al. 2022](#)) spectrum. For reference, the best-fit *Planck* 2018  $\Lambda\text{CDM}$  spectrum is plotted along side them. The middle panel shows the deviation from the *Planck*  $\Lambda\text{CDM}$  solution, in units of  $\sigma_\ell$  from each individual pipeline, while the bottom panel shows the fractional difference with respect to the *Planck*  $\Lambda\text{CDM}$  spectrum.

The spectrum has been computed following the same Gaussianized Blackwell-Rao (GBR) ([Rudjord et al. 2009](#)) analysis described in ([Colombo et al. 2022](#)). We performed a CMB temperature resampling by imposing a prior on the spacial correlations of the CMB, and masking the galactic plane with the BEYONDPLANCK temperature mask. The GBR estimator is computed using 500 CMB samples.

At  $\ell \lesssim 500$ , each of these datasets are signal-dominated, and each of the spectra agree at nearly every multipole. No-



**Fig. 25.** Posterior mean CMB COSMOGLOBE temperature map, smoothed to an angular resolution of 14' FWHM.



**Fig. 26.** Posterior mean CMB COSMOGLOBE maps and their standard deviation.

table exceptions include specifically the quadrupole, which we will discuss in Sect. 6.1.3, and  $50 \lesssim \ell \lesssim 100$ . In the latter, the COSMOGLOBE multipoles are consistently  $\geq 5\%$  higher than the  $\Lambda\text{CDM}$  solution.

At  $\ell \gtrsim 500$ , the WMAP9 solution becomes noise dominated, and there is much larger variation between the multipoles. While *Planck* and BEYONDPLANCK mostly agree with the COSMOGLOBE solution, the convergence properties of the power spectrum are much worse, and would require an order of magnitude more samples to converge (Colombo et al. 2022). We therefore caution against overinterpreting the specific values from the COSMOGLOBE power spectrum in this regime.

### 6.1.3. Low- $\ell$ anomalies

Even though the CMB is well described by  $\Lambda\text{CDM}$ , there are several anomalies, especially at low multipoles, that seem to be in tension with  $\Lambda\text{CDM}$ . As a rule, the existence of these anomalies is not debated, but rather their significance within  $\Lambda\text{CDM}$ . Indeed, as argued by Bennett et al. (2011), the existence of these anomalies may simply be described by the look-elsewhere effect. The relevant question is not “How unlikely is  $\Lambda\text{CDM}$  given this effect?” but rather “How unlikely is this effect given  $\Lambda\text{CDM}$ ?“.

In standard analysis of CMB anomalies, one computes the traditional maximum-likelihood maps of the CMB and generate a single value for each of the anomaly statistics. One then compares these values to simulations of the  $\Lambda\text{CDM}$  in a frequentist

**Table 7.** Comparison of Solar dipole measurements from *COBE*, *WMAP*, and *Planck*.

EXPERIMENT	AMPLITUDE [ $\mu\text{K}_{\text{CMB}}$ ]	GALACTIC COORDINATES		REFERENCE
		<i>l</i> [deg]	<i>b</i> [deg]	
<i>COBE</i> <sup>a,b</sup>	3358 $\pm$ 23	264.31 $\pm$ 0.16	48.05 $\pm$ 0.09	Lineweaver et al. (1996)
<i>WMAP</i> <sup>c</sup>	3355 $\pm$ 8	263.99 $\pm$ 0.14	48.26 $\pm$ 0.03	Hinshaw et al. (2009)
LFI 2015 <sup>b</sup>	3365.5 $\pm$ 3.0	264.01 $\pm$ 0.05	48.26 $\pm$ 0.02	Planck Collaboration II (2016)
HFI 2015 <sup>d</sup>	3364.29 $\pm$ 1.1	263.914 $\pm$ 0.013	48.265 $\pm$ 0.002	Planck Collaboration VIII (2016)
LFI 2018 <sup>b</sup>	3364.4 $\pm$ 3.1	263.998 $\pm$ 0.051	48.265 $\pm$ 0.015	Planck Collaboration II (2020)
HFI 2018 <sup>d</sup>	3362.08 $\pm$ 0.99	264.021 $\pm$ 0.011	48.253 $\pm$ 0.005	Planck Collaboration III (2020)
Bware	3361.90 $\pm$ 0.40	263.959 $\pm$ 0.019	48.260 $\pm$ 0.008	Delouis et al. (2021)
<i>Planck</i> PR4 <sup>a,c</sup>	3366.6 $\pm$ 2.6	263.986 $\pm$ 0.035	48.247 $\pm$ 0.023	Planck Collaboration Int. LVII (2020)
BEYONDPLANCK <sup>e</sup>	3362.7 $\pm$ 1.4	264.11 $\pm$ 0.07	48.279 $\pm$ 0.026	Colombo et al. (2022)
COSMOGLOBE <sup>e</sup>	<b>3366.2 <math>\pm</math> 1.4</b>	<b>264.08 <math>\pm</math> 0.07</b>	<b>48.273 <math>\pm</math> 0.024</b>	This work

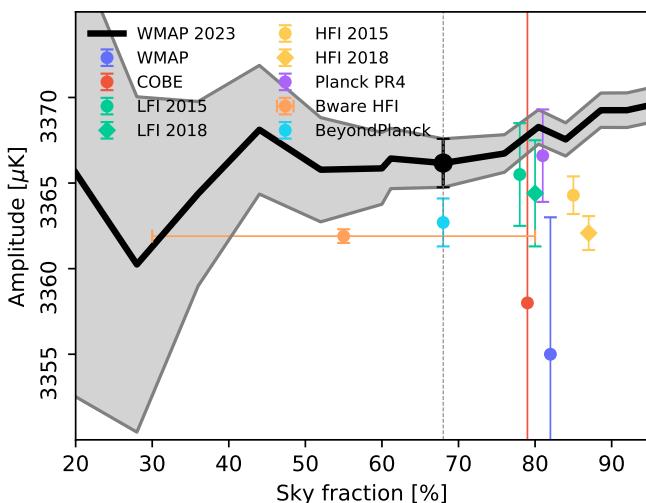
<sup>a</sup> Statistical and systematic uncertainty estimates are added in quadrature.

<sup>b</sup> Computed with a naive dipole estimator that does not account for higher-order CMB fluctuations.

<sup>c</sup> Computed with a Wiener-filter estimator that estimates, and marginalizes over, higher-order CMB fluctuations jointly with the dipole.

<sup>d</sup> Higher-order fluctuations as estimated by subtracting a dipole-adjusted CMB-fluctuation map from frequency maps prior to dipole evaluation.

<sup>e</sup> Estimated with a sky fraction of 68 %. Error bars include only statistical uncertainties, as defined by the global COSMOGLOBE posterior framework, and they thus account for instrumental noise, gain fluctuations, parametric foreground variations etc.



**Fig. 27.** CMB dipole amplitude as a function of sky fraction. The gray band indicates the 68 % posterior confidence region.

approach to see how rare it is for a random realization of the CMB to exhibit the anomalies.

In our case, we have a set of CMB maps outputted from the Gibbs chain, which allows us to create histograms of these statistics. This allows us to get full control over systematic uncertainties of the low- $\ell$  anomalies which is difficult in traditional approaches. This allows us to answer the question of whether the anomalies actually are in our CMB, or if previous measurements of the CMB simply have not accounted for all systematic effects.

This section will look at some of statistics of well-known the low- $\ell$  anomalies and compare our findings with the previous results in the BEYONDPLANCK analysis in Colombo et al. (2022). This will tell us how much more information is gained about these anomalies by including the *WMAP* dataset to the LFI of *Planck*.

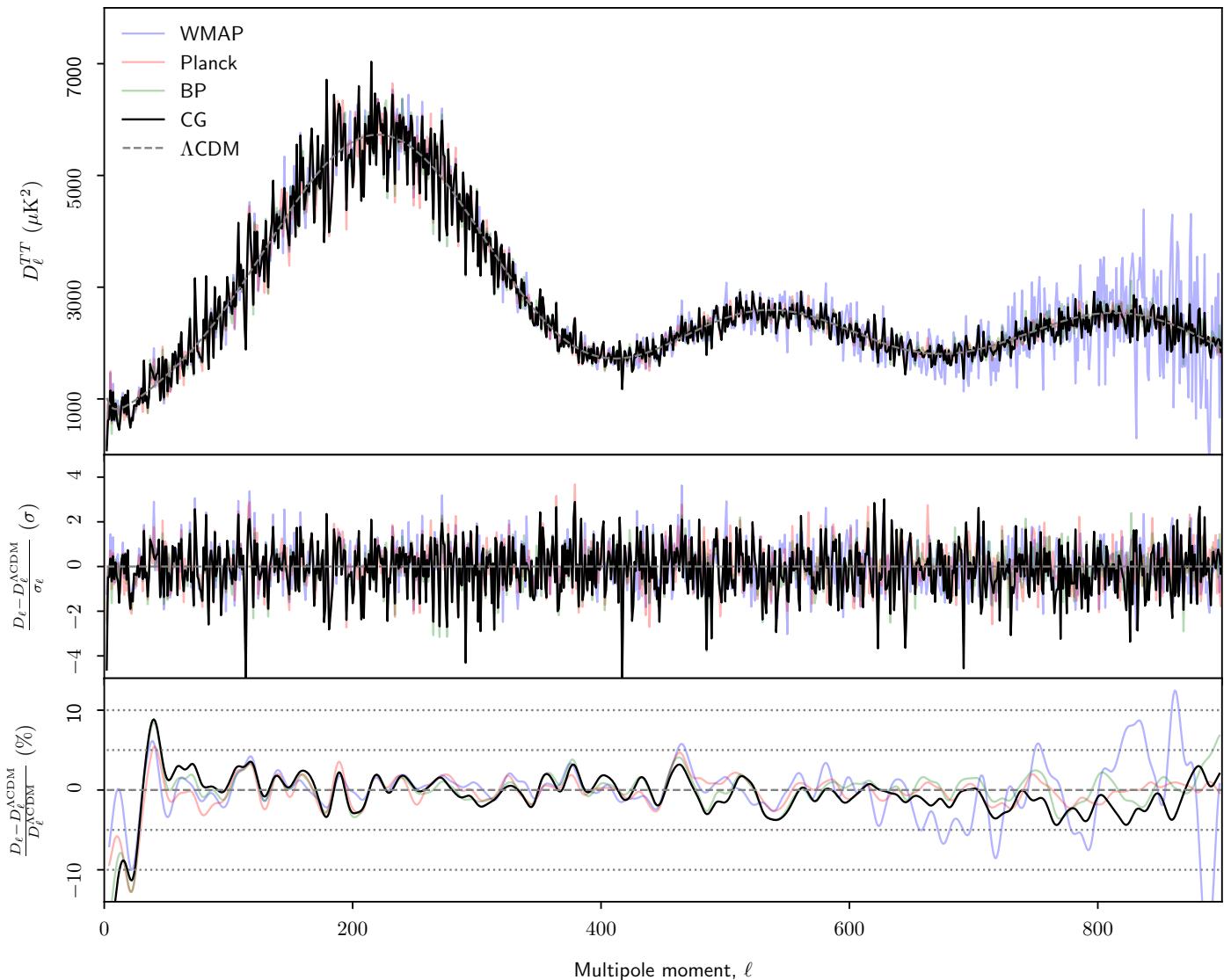
It has been noted since *COBE*-DMR that the quadrupole amplitude of our CMB is lower than expected from  $\Lambda\text{CDM}$  (Bennett et al. 1992). This has later been confirmed in *WMAP* (Hinshaw et al. 2003) and *Planck* (Planck Collaboration Int. XV 2014), but with large discrepancies in mean value and error bars.

The *WMAP* team found in the 7-year analysis a best-fit value of  $201 \mu\text{K}^2$  (Larson et al. 2011) which dropped to  $151 \mu\text{K}^2$  in the 9-year analysis (Hinshaw et al. 2013). The naive Fisher uncertainty on  $\sigma_2$  was reported by Hinshaw et al. (2013) to be  $9 \mu\text{K}^2$  which only accounted for a noise-only estimate. This error bar corresponds to a roughly  $5\sigma$  discrepancy.

*Planck* later found in 2013 and 2018  $\sigma_2$  to be  $299$  and  $226 \mu\text{K}^2$ , respectively, with a naive Fisher uncertainty yielded a  $8\sigma$  discrepancies. This shows that measuring  $\sigma_2$  is difficult and that instrumental white-noise is not the dominant source of uncertainty.

With the set of COSMOGLOBE maps of the CMB we can quantify the marginal posterior distribution  $P(\sigma_2 | \mathbf{d})$ , and we show the posterior distribution represented as a Gaussian with  $\sigma_2 = 132 \pm 71 \mu\text{K}^2$  in Fig. 29. Despite the fact that COSMOGLOBE also incorporates the LFI dataset, this is an 8-fold of the *WMAP* Fisher uncertainty. This is a clear demonstration of why one must take instrumental and astrophysical parameters into account when measuring  $\sigma_2$ . Fig. 29 compares the COSMOGLOBE results with BEYONDPLANCK which found  $\sigma_2 = 229 \pm 97 \mu\text{K}^2$ , showing that the inclusion of *WMAP* significantly lowers the measured value of  $\sigma_2$ . We also show  $10^5$  realizations of  $\sigma_2$  given *Planck* 2018 best-fit  $C_2^{\Lambda\text{CDM}} = 1064.6 \mu\text{K}^2$ .

To better understand how anomalous  $\sigma_2$  is, we can ask what the probability of  $C_2$  be large given our set of measured  $\sigma_2$ . This can be done by applying the Blackwell-Rao estimator described by Chu et al. (2005). We map out the marginal posterior distribution  $P(C_2 | \sigma_2)$  as a function of  $C_2$  and mark where  $C_2^{\Lambda\text{CDM}} = 1064.6 \mu\text{K}^2$  by a grey vertical line. We show this in Fig. 30 find that the probability for  $C_2$  to exceed  $C_2^{\Lambda\text{CDM}}$  to be 11.1 % for COSMOGLOBE (LFI + *WMAP*) and 21.7 % for BEYONDPLANCK (LFI) shown as shaded areas.



**Fig. 28.** (Top:) Angular CMB temperature power spectrum,  $D_\ell^{TT}$ , as derived by COSMOGLOBE (black), BEYONDPLANCK (green), *Planck* (red), and *WMAP9* (blue). The best-fit *Planck* 2018  $\Lambda\text{CDM}$  spectrum is showed in dashed gray. (Middle:) Residual power spectrum relative to  $\Lambda\text{CDM}$ , measured relative to the quoted error bars,  $(D_\ell - D_\ell^{\Lambda\text{CDM}})/\sigma_\ell$ . For pipelines that report asymmetric error bars,  $\sigma_\ell$  is taken to be the average of the upper and lower error bar. (Bottom:) Fractional difference with respect to the *Planck*  $\Lambda\text{CDM}$  spectrum. In this panel, each curve has been boxcar averaged with a window of  $\Delta\ell = 100$  to suppress random fluctuations.

## 6.2. Galactic foregrounds

As described in Sect. 2.4, we adopt a similar sky model to that of [BeyondPlanck \(2022\)](#). Explicitly, in this work the low frequency component (free-free, anomalous microwave, and synchrotron) emission amplitudes and their spectral parameters are fit in total intensity, which are described below. The thermal dust amplitude is also fit here, though the addition of *Planck* 857 GHz provides most of the constraining power. The solution to the thermal dust amplitude is shown in the bottom left panel of Fig. 34.

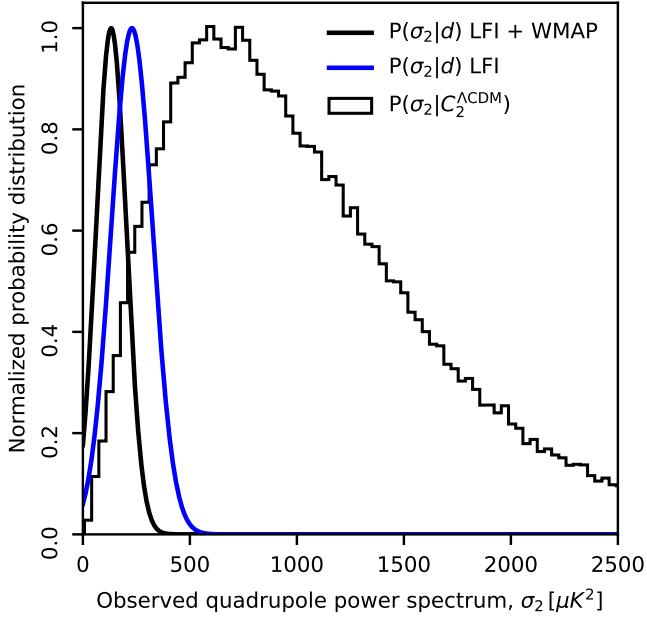
### 6.2.1. Free-free

In the top left panel of Fig. 34, we see the free-free component, which shows good agreement with previous full sky component separation studies ([Planck Collaboration X 2016](#); [Andersen et al. 2022](#)). We note that compared to the [Planck Collaboration X \(2016\)](#) analysis, there is less diffuse structure in the free-free component, which is driven by the imposition of a prior at high

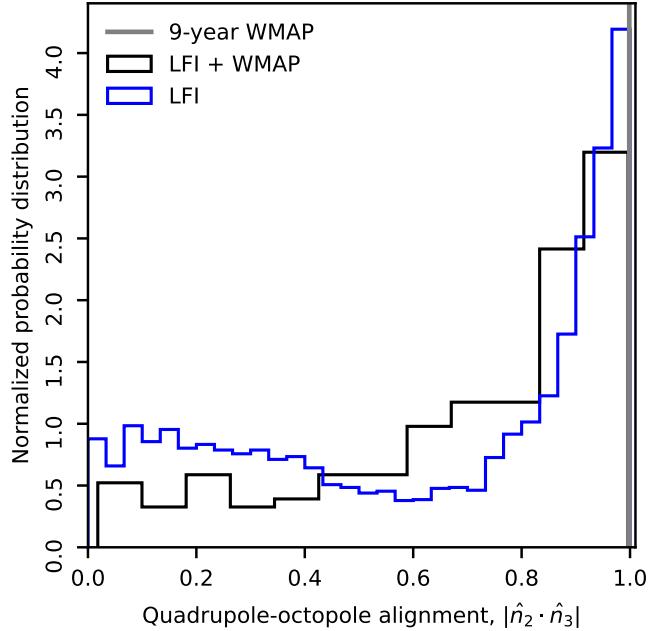
Galactic latitudes ([Andersen et al. 2022](#)). However, in high emission regions, such as the Galactic plane and the Gum Nebula, we see strong agreement.

### 6.2.2. Anomalous microwave emission

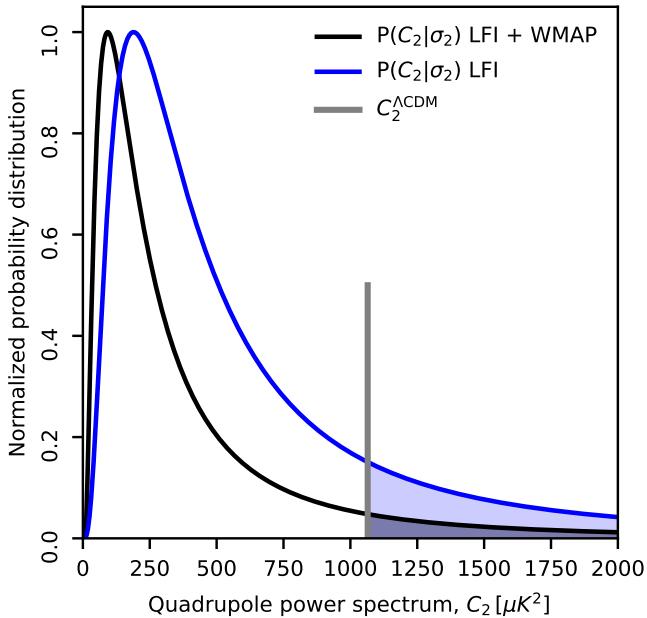
For the AME, we see a differing morphology from both [Planck Collaboration X \(2016\)](#) and [Andersen et al. \(2022\)](#). The most notable difference is the lack of extended diffuse structure in this work, with a marginal shift in the overall direction of the component's dipole. This is due to the degeneracy between the  $K$ -band gain and the AME dipole, as described at length in Sect. 7.3. Both this analysis and [Andersen et al. \(2022\)](#) differ from the [Planck Collaboration X \(2016\)](#) AME solution by showing less extended diffuse structure, and most visibly notable is the  $\rho$ -Ophiuchi complex, which appears as a hole in the AME component in this work. Determination of the AME SED and a more direct analysis of the AME component will be discussed in [Watts et al. \(2023\)](#).



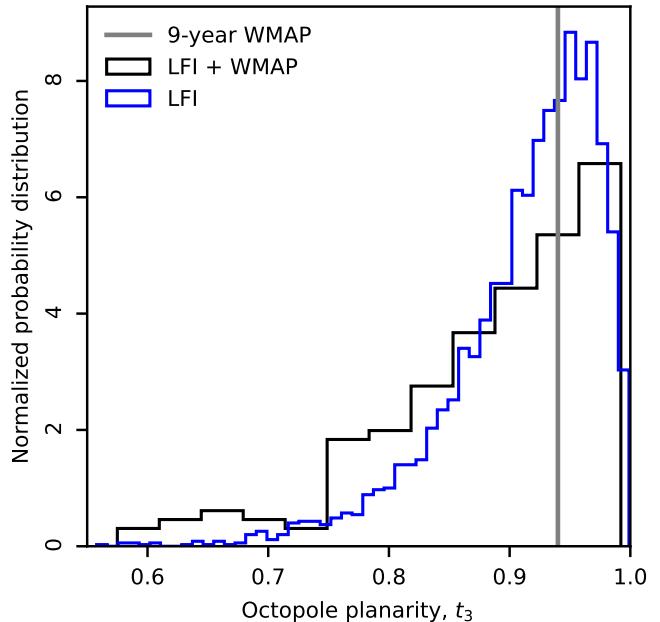
**Fig. 29.** Histogram of 100 000 realizations of potential realizations  $\sigma_2$  given  $C_2^{\Lambda\text{CDM}} = 1064.7 \mu\text{K}^2$  compared with the measured power spectrum  $\sigma_2$  of our universe for COSMOGLOBE (black) and BEYONDPLANCK (blue).



**Fig. 31.** The quadrupole-octopole alignment of COSMOGLOBE compared with BEYONDPLANCK.



**Fig. 30.** Marginal probability distribution of the ensemble-averaged  $C_2$  given the data,  $P(C_2 | d)$ , as measured by COSMOGLOBE (black) and BEYONDPLANCK (blue).

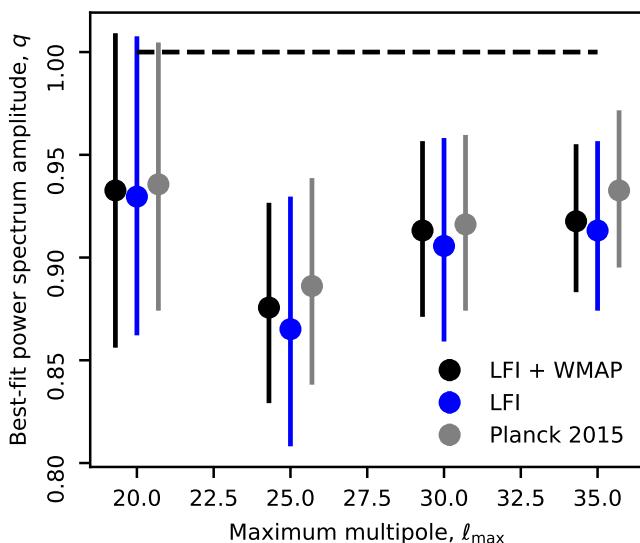


**Fig. 32.** The octopole planarity statistics  $t_3$  compared with the BEYONDPLANCK analysis (blue).

### 6.2.3. Synchrotron

In total intensity, the reprocessed fullsky Haslam map (Re-mazeilles et al. 2015) at 408 MHz is used as an anchor for the full sky synchrotron emission. As such, the component separation solution shown in the bottom right panel of Fig. 34 clearly shares similar morphology to both the *Planck* and BEYONDPLANCK analyses, with some deviations around point sources, which are explicitly modeled in this work.

The Stokes  $Q$  and  $U$  synchrotron amplitude mean and standard deviations are shown in Fig. 35. The mean amplitude map is in good agreement with that presented in Svalheim et al. (2022b), though the morphology of the standard deviation map matches more closely with that presented in Herman et al. (2022), which shows convolution of the WMAP and *Planck* scanning strategies. However, the uncertainty of the signal along the Galactic plane is more tightly constrained in this work as the AME component is assumed to be unpolarized.



**Fig. 33.** Best-fit amplitude,  $q$ , of the low multipole power spectrum  $C_\ell = qC_\ell^{\Lambda\text{CDM}}$ ,  $2 \leq \ell \leq \ell_{\max}$  compared to *Planck* 2015 (grey) and BeyondPlanck (blue).

#### 6.2.4. $\chi^2$ and map space residuals

The quality of the component separation procedure is evaluated through a reduced- $\chi^2$  map, shown in Fig. 36, as well as through the map space residuals Fig. B.6. Figure 36 shows a sum of all three Stokes parameters, and we can see, morphologically, that the poorly fit parts of the sky are driven by structures associated with features in total intensity. The poorly fit pixels are generally constrained to the Galactic plane, mostly at the Galactic center, though we note some poor fits in regions such as the Orion region,  $\rho$  Ophiucus, and Large Magellanic Cloud. Future work to find a more complete model of AME will likely help clean up this picture.

#### 6.3. WMAP-versus-LFI signal-to-noise ratio comparison

### 7. Outstanding issues

As shown in the previous sections, there are very few residuals, artifacts, or systematics that make this jointly processed dataset, hereafter referred to as COSMOGLOBE data release 1 (CG1). However, the global nature of this analysis allows us to identify issues in the data processing that will otherwise have gone unnoticed. In this section, we enumerate the issues we have encountered in CG1, and which we plan to improve upon in future data releases.

#### 7.1. Noise modeling

As demonstrated in Sect. 4.4, the  $\chi^2$  per TOD scan was discrepant up to the  $10\sigma$  level. The main driver of this model failure is in the noise modeling, and in particular the incompleteness of a power law correlated noise model. In practice, this is due to the algorithmic choice to not fit the white noise level as a parameter on the same footing as all other noise parameters. Inspecting Fig. 9 shows that the high-frequency noise is essentially fixed to the noise PSD at the sampling frequency. If  $\sigma_0$  were a free parameter in this particular parametric fit, it would be driven by the intermediate frequencies 2–6 Hz at the expense of a good fit at the highest frequencies.

The particular case of W413’s PSD is a noise spectrum that could easily be modeled as a spectrum that is continuing to drop beyond the sampling rate, not dissimilar to the two-pole Bessel filter implemented in *WMAP*’s electronics (Jarosik et al. 2003b). In practice, the white noise can be identified with the flat portion of the spectrum well above  $f_{\text{knee}}$ , but in the case of these noise spectra, there is no such flat portion, challenging the very existence of “white noise” for this particular radiometer. Additionally, a Bessel filter tail could affect the signal band as well, requiring more detailed modeling of the noise.

In practice, the decomposition of instrumental noise into a “white” component and a correlated component is very useful, and provides a stringent test for the final data products. Indeed, the particular model failure was so subtle that such a description of noise being split into scale-dependent and scale-independent would have made it nearly impossible to detect such an issue. For the case of *WMAP* data, there is a natural need to improve the noise PSD modeling, especially when a successful parameterization was found by the *WMAP* team in time space. In practice, this will likely be useful for the analysis of other CMB experiments, and will be of broad use in the future.

#### 7.2. V- and W-band quadrupole residual

As mentioned in Sect. 5.2, there is a  $2\mu\text{K}$  quadrupolar residual in the V-band and W-band residuals that is aligned with the ecliptic pole, and roughly perpendicular to the Solar dipole. As this quadrupole does not exist in lower DAs, and V- and W-bands have gains that are consistent with *WMAP*’s at the 1 % level, it is more likely that there is a subtle error in the data processing.

In an early stage of this analysis, a large quadrupolar signal was induced due to an error in the orbital dipole calculation. Essentially, a single satellite velocity was assumed for an entire scan, which proved to be a poor approximation over  $\sim 3$ -day period. A linear interpolation between scans fixed this issue, and a cubic interpolation provided a negligible improvement.

This points generally to long-timescale effects causing quadrupolar signals. A similar issue, an  $8\mu\text{K}$  dipole, was discovered by Jarosik et al. (2007), and determined to be due to an inadequacy in the gain model. As mentioned earlier, we assume a linear baseline trend throughout a given scan, and allow correlated noise residuals to pick up longer scale fluctuations. Compared to the *WMAP* team’s approach of fitting cubic polynomials every hour, there is much more room for unmodeled temporal variation in zero-level. As the gain, correlated noise, and baseline are all deeply correlated, a subtle error in the baseline determination could easily induce a small quadrupolar signal.

#### 7.3. Degeneracy between K-band calibration and AME dipole

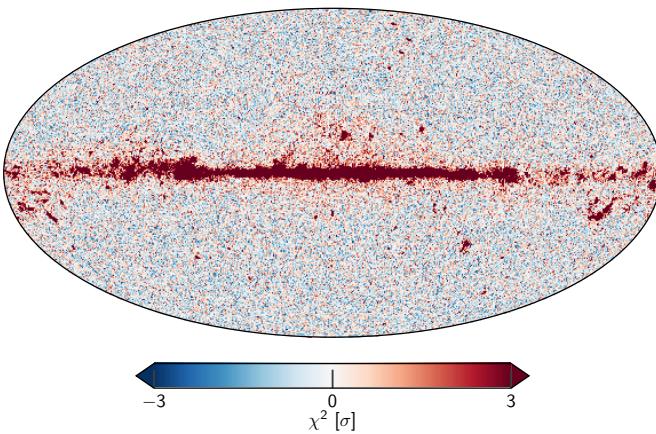
As discussed in Sect. 2.5, there is a strong degeneracy between *K*-band’s absolute calibration and the AME dipole that requires external information to break. In this work, we implemented a prior on the absolute calibration based on the effect it had on the best-fit AME dipole. In the posterior distribution, in which we show a slice of in Fig. 38, we can specifically compare *K*-band’s absolute calibration with the AME dipole values. Here the degeneracy between  $g_0$  and  $a_{1,\pm 1}^{\text{AME}}$  is quite apparent. Because there is no causal connection between *K*-band’s absolute gain and the AME dipole, a prior was clearly necessary for the analysis in this work to return sensible results.



**Fig. 34.** Foreground intensity maps, evaluated at their respective reference frequencies. (Top left) Free-free emission at 40 GHz. (Top right) Anomalous microwave emission evaluated at 22 GHz. (Bottom left) Thermal dust emission at 70 GHz. (Bottom right) Synchrotron emission evaluated at 408 MHz.



**Fig. 35.** Polarized synchrotron maps and their standard deviations evaluated at 30 GHz.



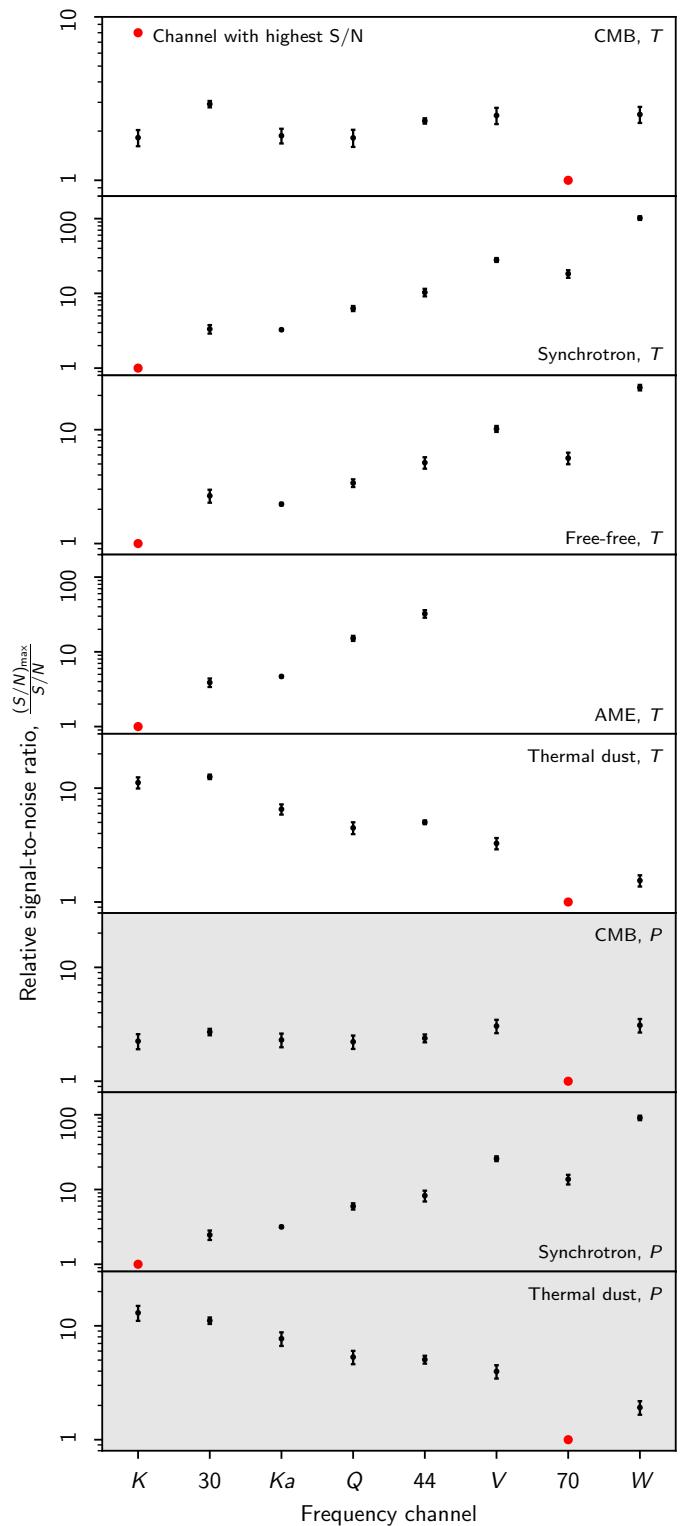
**Fig. 36.** Reduced- $\chi^2$ , using  $n_{\text{dof}} = 300$ , which comes from fitting to the regions outside of the  $K$ -band processing mask.

In the official *WMAP* pipeline, the degeneracy was effectively broken by using a preliminary  $K$ -band sky map and removing it from the timestream. In practice, both solutions are the result of scientific intuition solving an algorithmic issue. The COSMOGLOBE approach of using a prior on  $g_0$  comes from the strong prior that Galactic emission should not have a dipole aligned with the CMB's Solar dipole. The *WMAP* team's approach of using a previous iteration's map as a sky model comes from the strong prior that errors in the first iteration of the sky map are uncorrelated with the orbital dipole in the timestream.

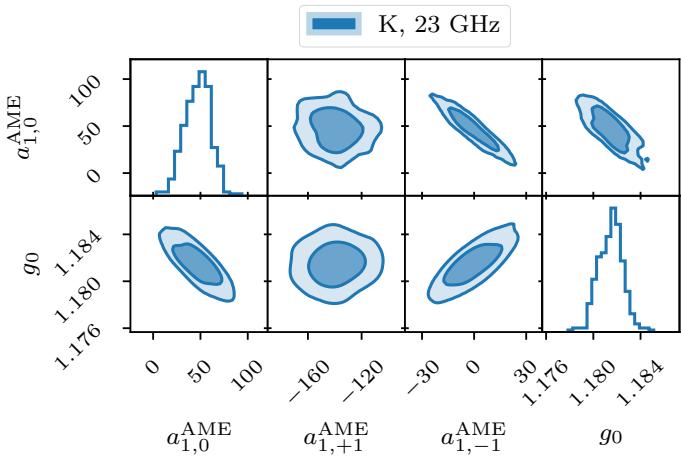
While the COSMOGLOBE  $K$ -band gain solution is driven by our prior belief in the size of the AME dipole, the overall instrumental solution generates maps that are consistent with our sky model at the  $1 \mu\text{K}$  level at high Galactic latitudes. Conversely, the *WMAP9* solution does not rely on any knowledge of the sky, but as a result induces poorly measured modes with amplitude  $2.5 \mu\text{K}$ .

Regardless of the details, an accurate model of the sky as observed by  $K$ -band is a necessary condition to obtaining an accurate measurement of the gain. The difficulty of obtaining an accurate AME model is of course compounded by the fact that the AME is brighter in  $K$ -band than any of the *WMAP* or *Planck* bands. This may be mitigated in the near future, following a joint *WMAP+LFI+QUIJOTE* analysis, but this of course depends on the signal-to-noise of AME in QUIJOTE's frequencies, and is further hindered by QUIJOTE's partial sky coverage.

A future analysis involving the most robust parts of the *WMAP9* and COSMOGLOBE analysis also has the potential to solve the  $g_0$ -ALM degeneracy. In particular, the COSMOGLOBE analysis did not directly use the housekeeping data to estimate the gain model. There is no a priori reason that the parameters in Eq. (48) cannot be included in the Gibbs chain. This would of course require detailed knowledge of the *WMAP* satellite's hardware only held by the core *WMAP* team. We hope that a joint effort between the *WMAP* team and COSMOGLOBE will help to solve this outstanding issue.



**Fig. 37.** Relative signal-to-noise ratios for WMAP and LFI channels and various components.



**Fig. 38.** Correlation between  $K$ -band's absolute calibration  $g_0$  [ $\text{du mK}^{-1}$ ] and AME dipole's spherical harmonic coefficients [ $\mu\text{K}_\text{RJ}$ ].

#### 7.4. Other minor effects

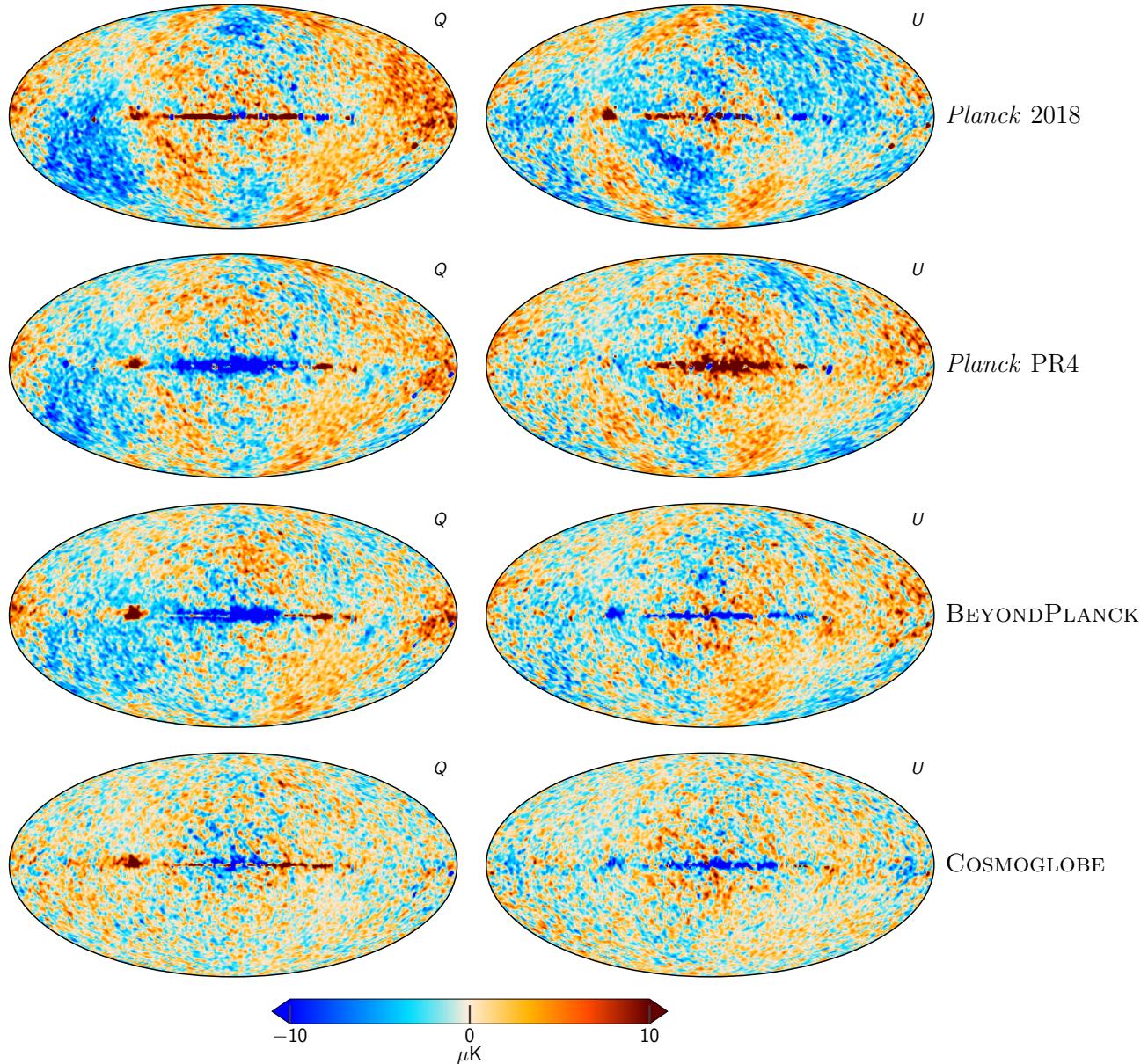
##### 7.4.1. Time-variable bandpass modeling

##### 7.4.2. Polarized sidelobe modeling

## 8. Conclusions

## References

- Ali-Haïmoud, Y. 2010, Astrophysics Source Code Library [ascl:1010.016]
- Ali-Haïmoud, Y., Hirata, C. M., & Dickinson, C. 2009, MNRAS, 395, 1055
- Alonso, D., Sanchez, J., Slosar, A., & LSST Dark Energy Science Collaboration. 2019, MNRAS, 484, 4127
- Andersen et al. 2022, A&A, in press [arXiv:2201.08188]
- Barnes, C., Hill, R. S., Hinshaw, G., et al. 2003, ApJS, 148, 51
- Barrett, R., Berry, M. W., Chan, T. F., et al. 1994, Templates for the Solution of Linear Systems (Society for Industrial and Applied Mathematics)
- Basyrov et al. 2022, A&A, submitted [arXiv:2208.14293]
- Bennett, C. L., Bay, M., Halpern, M., et al. 2003a, ApJ, 583, 1
- Bennett, C. L., Halpern, M., Hinshaw, G., et al. 2003b, ApJS, 148, 1
- Bennett, C. L., Hill, R. S., Hinshaw, G., et al. 2011, ApJS, 192, 17
- Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20
- Bennett, C. L., Smoot, G. F., Hinshaw, G., et al. 1992, ApJ, 396, L7
- BeyondPlanck. 2022, A&A, submitted [arXiv:2011.05609]
- Chu, M., Eriksen, H. K., Knox, L., et al. 2005, Phys. Rev. D, 71, 103002
- Colombo et al. 2022, A&A, submitted [arXiv:2208.14276]
- Delouis, J. M., Puget, J. L., & Vibert, L. 2021, A&A, 650, A82
- Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T. 1965, ApJ, 142, 414
- Dickinson, C., Davies, R. D., & Davis, R. J. 2003, MNRAS, 341, 369
- Frigo, M. & Johnson, S. G. 2005, Proceedings of the IEEE, 93, 216, special issue on "Program Generation, Optimization, and Platform Adaptation"
- Galloway et al. 2022a, A&A, in press [arXiv:2201.03509]
- Galloway et al. 2022b, A&A, in press [arXiv:2201.03478]
- Gjørøe et al. 2022, A&A, submitted [arXiv:2011.08082]
- Greason, M. R., Limon, M., Wollack, E., et al. 2012, Nine-Year Explanatory Supplement, 5th edn., Greenbelt, MD: NASA/GSFC
- Hensley, B., Murphy, E., & Staguhn, J. 2015, MNRAS, 449, 809
- Herman et al. 2022, A&A, in press [arXiv:2201.03530]
- Hill, R. S., Weiland, J. L., Odegard, N., et al. 2009, ApJS, 180, 246
- Hinshaw, G., Barnes, C., Bennett, C. L., et al. 2003, ApJS, 148, 63
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
- Hinshaw, G., Nolta, M. R., Bennett, C. L., et al. 2007, ApJS, 170, 288
- Hinshaw, G., Weiland, J. L., Hill, R. S., et al. 2009, ApJS, 180, 225
- Hu, W., Fukugita, M., Zaldarriaga, M., & Tegmark, M. 2001, ApJ, 549, 669
- Ihle et al. 2022, A&A, in press [arXiv:2011.06650]
- Jarosik, N., Barnes, C., Bennett, C. L., et al. 2003a, ApJS, 148, 29
- Jarosik, N., Barnes, C., Greason, M. R., et al. 2007, ApJS, 170, 263
- Jarosik, N., Bennett, C. L., Dunkley, J., et al. 2011, ApJS, 192, 14
- Jarosik, N., Bennett, C. L., Halpern, M., et al. 2003b, ApJS, 145, 413
- Lange, A. E., Ade, P. A., Bock, J. J., et al. 2001, Phys. Rev. D, 63, 042001
- Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, ApJS, 192, 16
- Larson, D., Weiland, J. L., Hinshaw, G., & Bennett, C. L. 2015, ApJ, 801, 9
- Lineweaver, C. H., Tenorio, L., Smoot, G. F., et al. 1996, ApJ, 470, 38
- Mather, J. C., Cheng, E. S., Cottingham, D. A., et al. 1994, ApJ, 420, 439
- Mather, J. C., Fixsen, D. J., Shafer, R. A., Mosier, C., & Wilkinson, D. T. 1999, ApJ, 512, 511
- Notari, A. & Quartin, M. 2015, Journal of Cosmology and Astroparticle Physics, 2015, 047–047
- Page, L., Jackson, C., Barnes, C., et al. 2003, ApJ, 585, 566
- Paradiso et al. 2022, A&A, submitted [arXiv:2205.10104]
- Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419
- Planck Collaboration II. 2016, A&A, 594, A2
- Planck Collaboration VIII. 2016, A&A, 594, A8
- Planck Collaboration IX. 2016, A&A, 594, A9
- Planck Collaboration X. 2016, A&A, 594, A10
- Planck Collaboration I. 2020, A&A, 641, A1
- Planck Collaboration II. 2020, A&A, 641, A2
- Planck Collaboration III. 2020, A&A, 641, A3
- Planck Collaboration IV. 2018, A&A, 641, A4
- Planck Collaboration V. 2020, A&A, 641, A5
- Planck Collaboration Int. XV. 2014, A&A, 565, A103
- Planck Collaboration Int. XLVI. 2016, A&A, 596, A107
- Planck Collaboration Int. LVII. 2020, A&A, 643, A42
- Planck Collaboration LVII. 2020, A&A, 643, A42
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical Recipes 3rd Edition: The Art of Scientific Computing, 3rd edn. (USA: Cambridge University Press)
- Remazeilles, M., Dickinson, C., Banday, A. J., Bigot-Sazy, M.-A., & Ghosh, T. 2015, MNRAS, 451, 4311
- Rubino-Martín, J. A., Guidi, F., Génova-Santos, R. T., et al. 2023, MNRAS, 519, 3383
- Rudjord, Ø., Groeneboom, N. E., Eriksen, H. K., et al. 2009, ApJ, 692, 1669
- Shewchuk, J. R. 1994, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Edition 1 $\frac{1}{4}$ , <http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf>
- Silsbee, K., Ali-Haïmoud, Y., & Hirata, C. M. 2011, MNRAS, 411, 2750
- Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJ, 396, L1
- Stevenson, M. A. 2014, The Astrophysical Journal, 781, 113
- Svalheim et al. 2022a, A&A, in press [arXiv:2201.03417]
- Svalheim et al. 2022b, A&A, in press [arXiv:2011.08503]
- Thommesen, H., Andersen, K. J., Aurlien, R., et al. 2020, A&A, 643, A179
- van der Vorst, H. A. 1992, SIAM Journal on Scientific and Statistical Computing, 13, 631
- Watts et al. 2022, A&A, in press [arXiv:2202.11979]
- Watts et al. 2023, A&A, in preparation [arXiv:20xx.xxxxx]
- Weiland, J. L., Odegard, N., Hill, R. S., et al. 2011, ApJS, 192, 19



**Fig. 39.** Difference maps between the *Planck* 30 GHz and *WMAP* K-band maps. The columns are (1) *Planck* 2018 v. *WMAP*9, (2) *Planck* PR4 v. *WMAP*9, (3) BEYONDPLANCK v. *WMAP*9, and (4) COSMOGLOBE *Planck* 30 GHz and *WMAP* K-band both produced in this paper. All maps have been smoothed to a common resolution of  $2^\circ$  FWHM, and the K-band map has been scaled by 0.495 to account for different central frequencies, assuming a synchrotron spectral index  $\beta_s = -3.1$ .

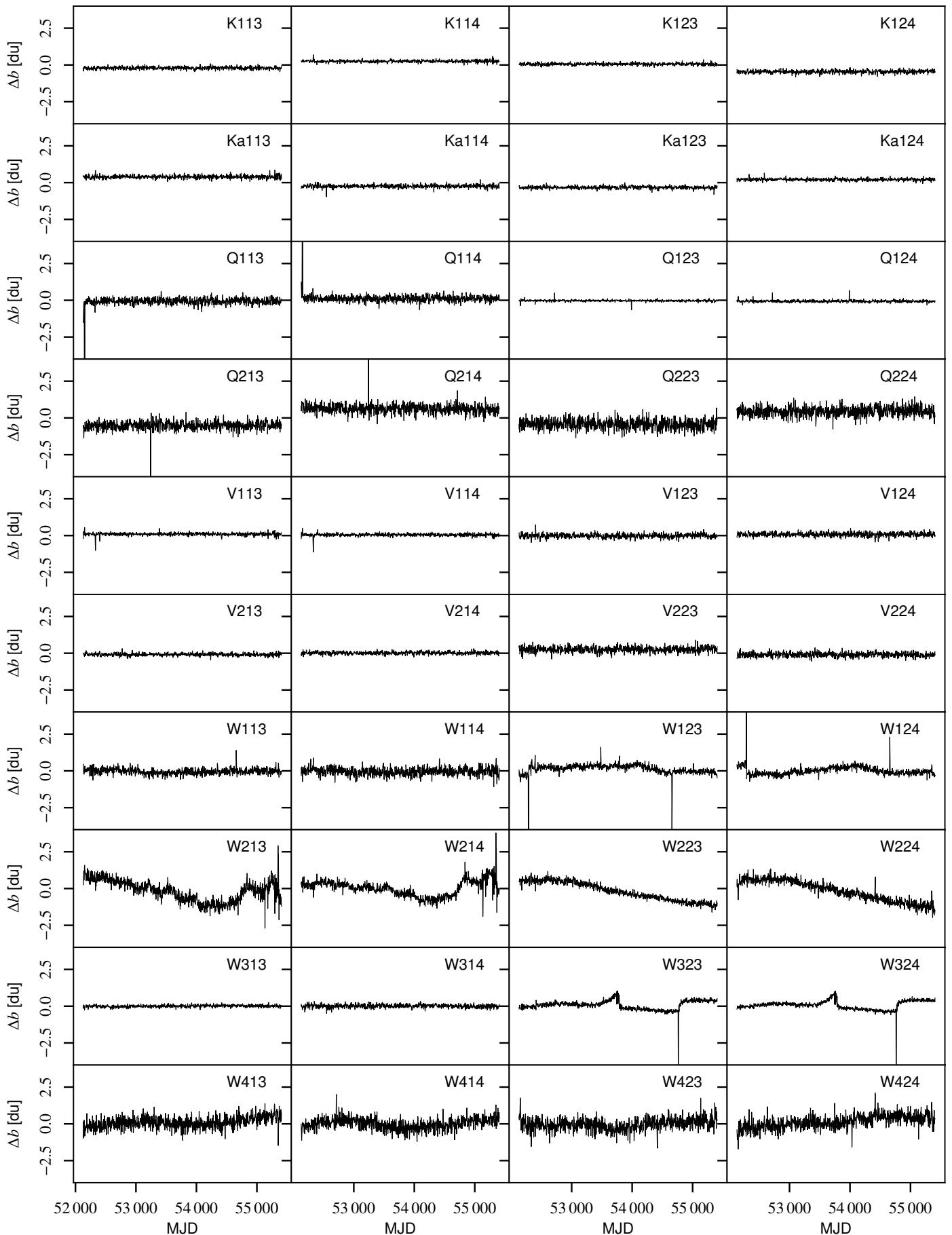
## **Appendix A: Survey of instrumental parameters**

*Appendix A.1: Gain, baselines, noise and  $\chi^2$*

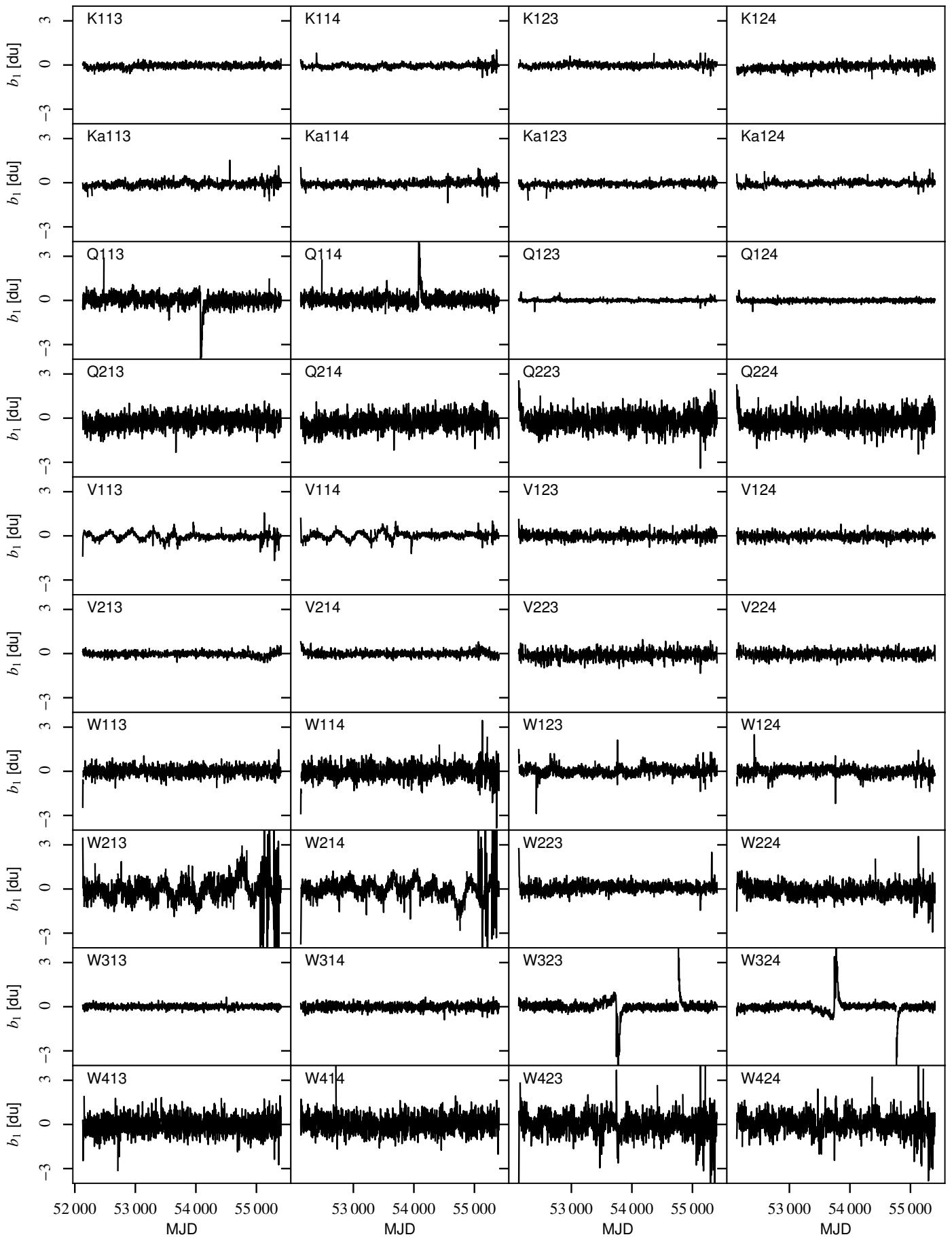
*Appendix A.2: Transmission imbalance*

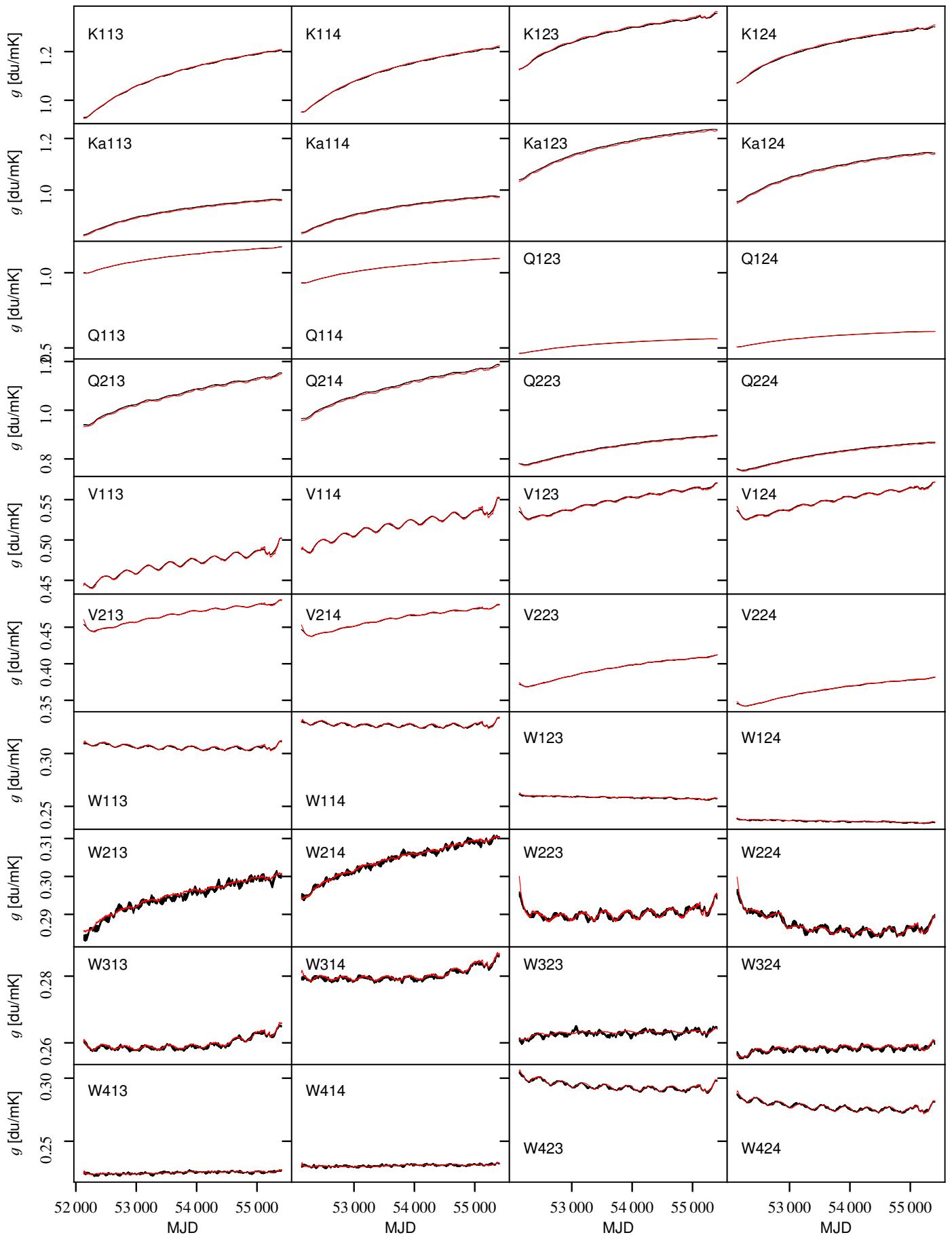
## **Appendix B: *WMAP* frequency map survey**

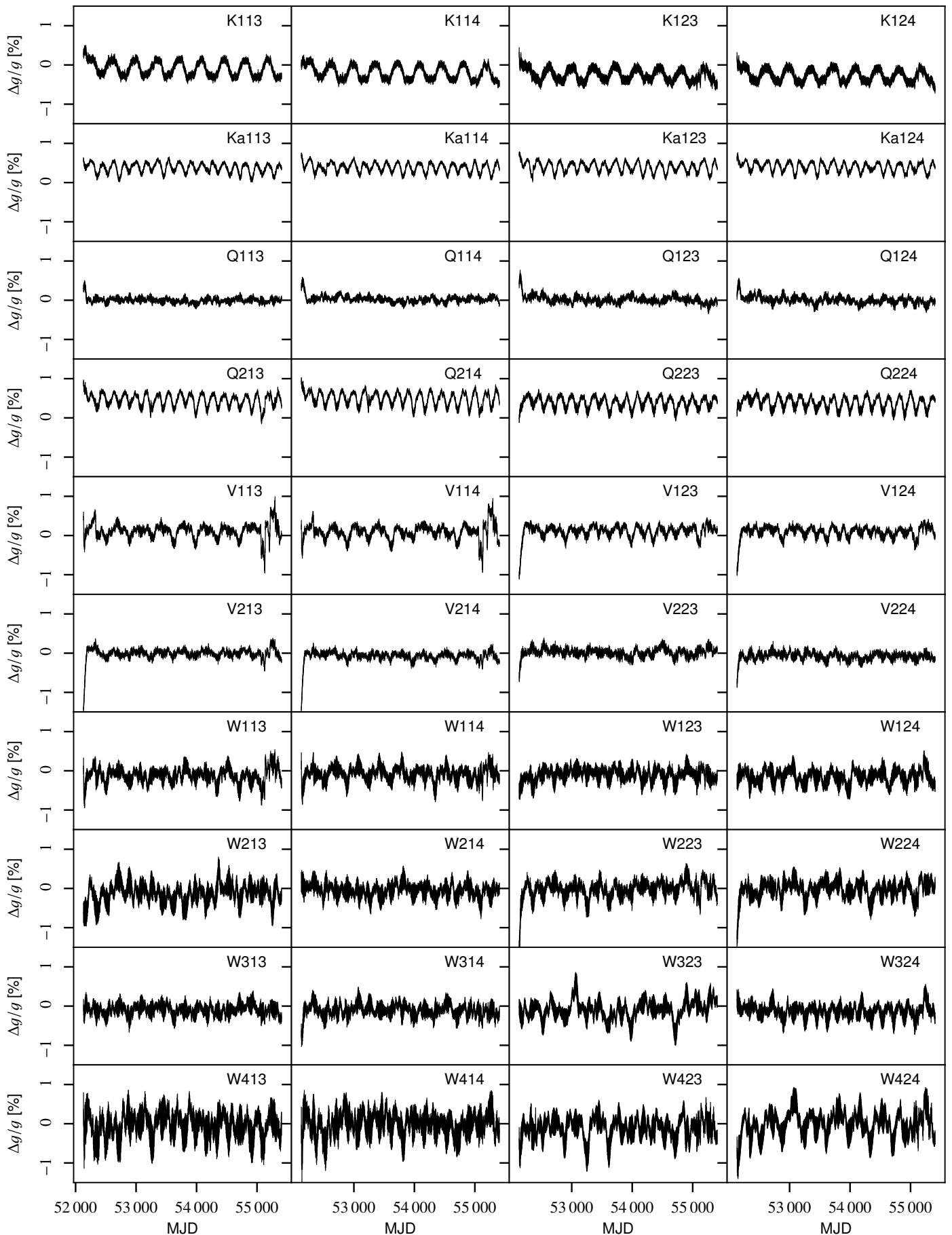
## **Appendix C: Comparison with BEYONDPLANCK LFI results**

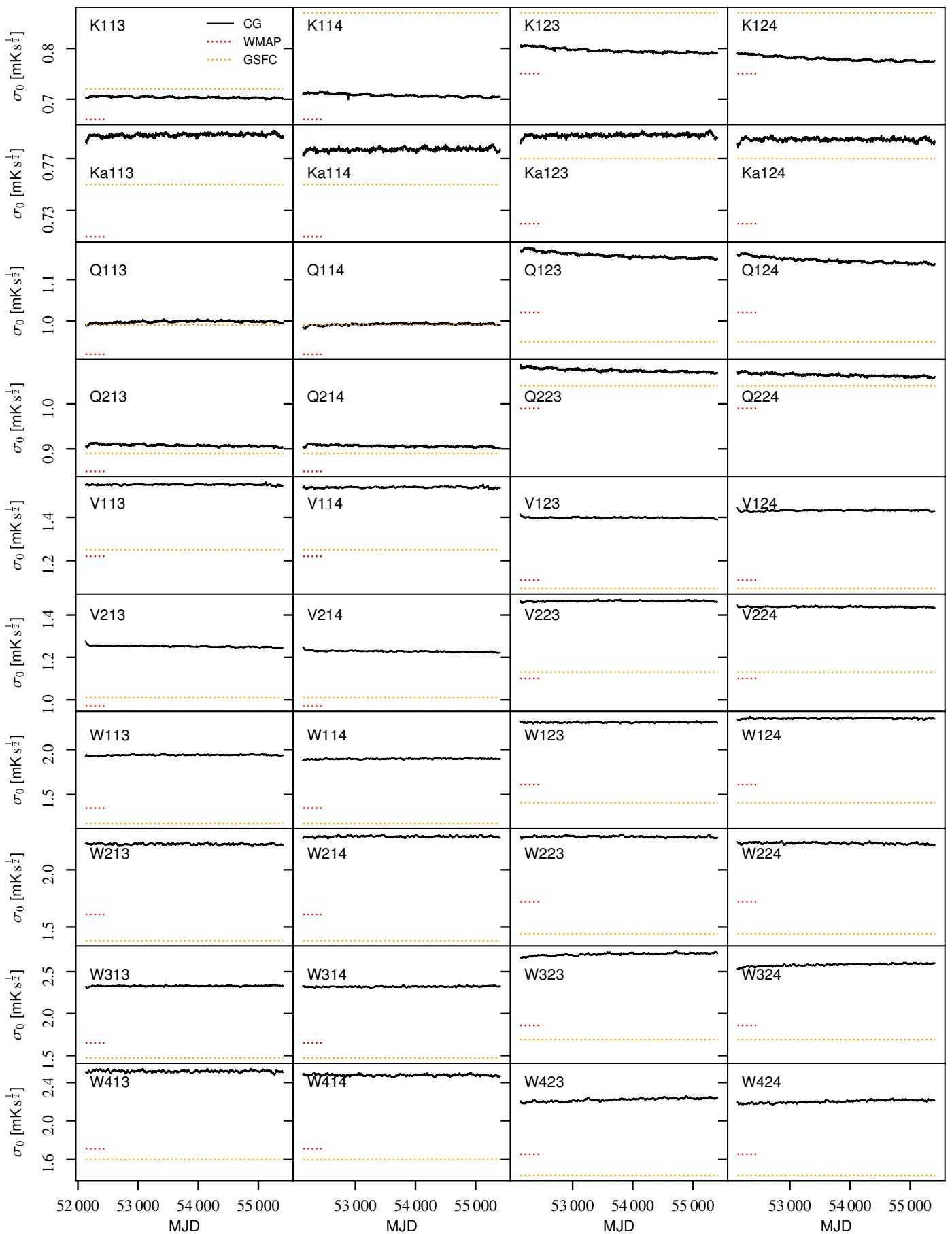


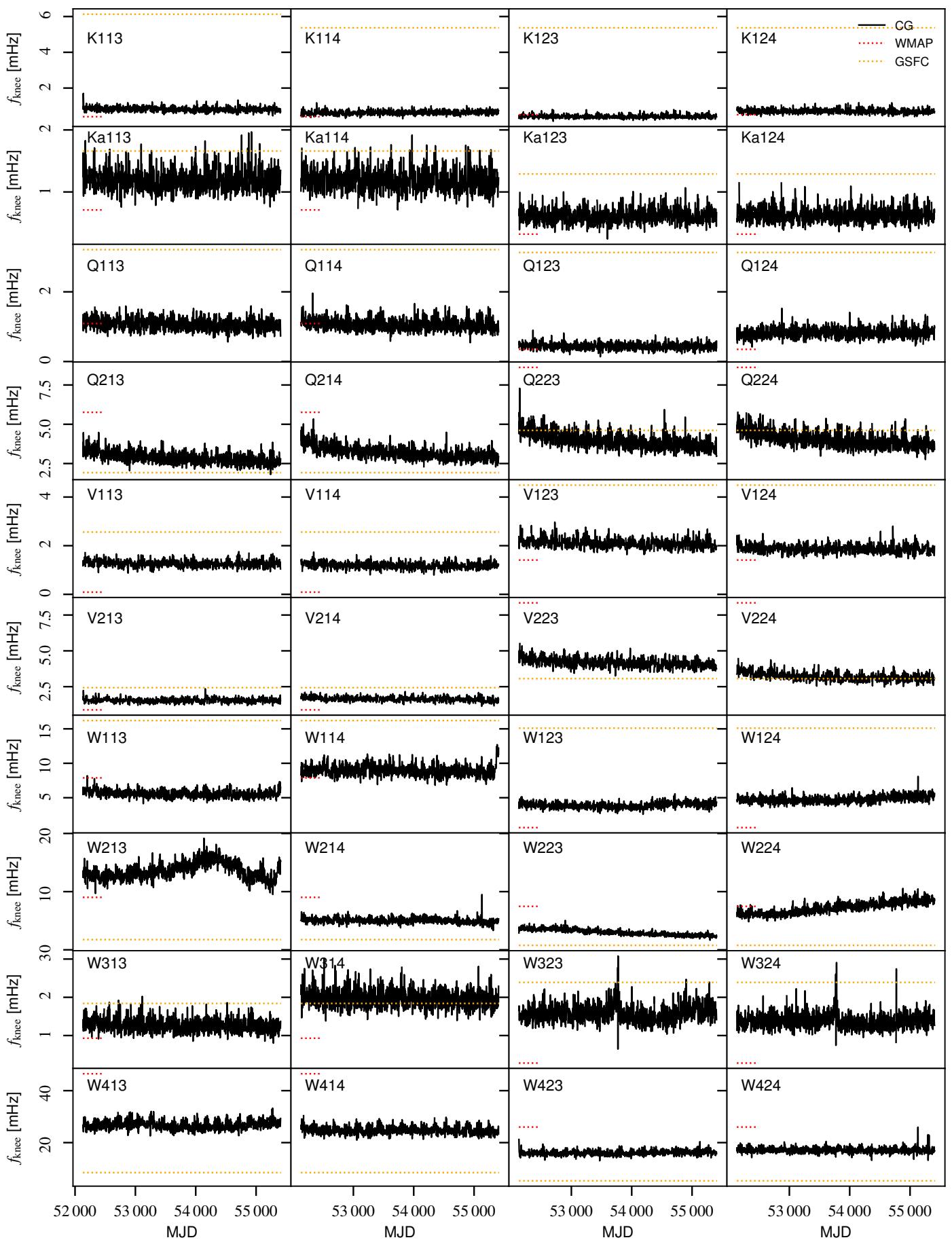
**Fig. A.1.** Difference in baseline solution,  $b_0^{\text{CG}} - b_0^{\text{WMAP}}$ .

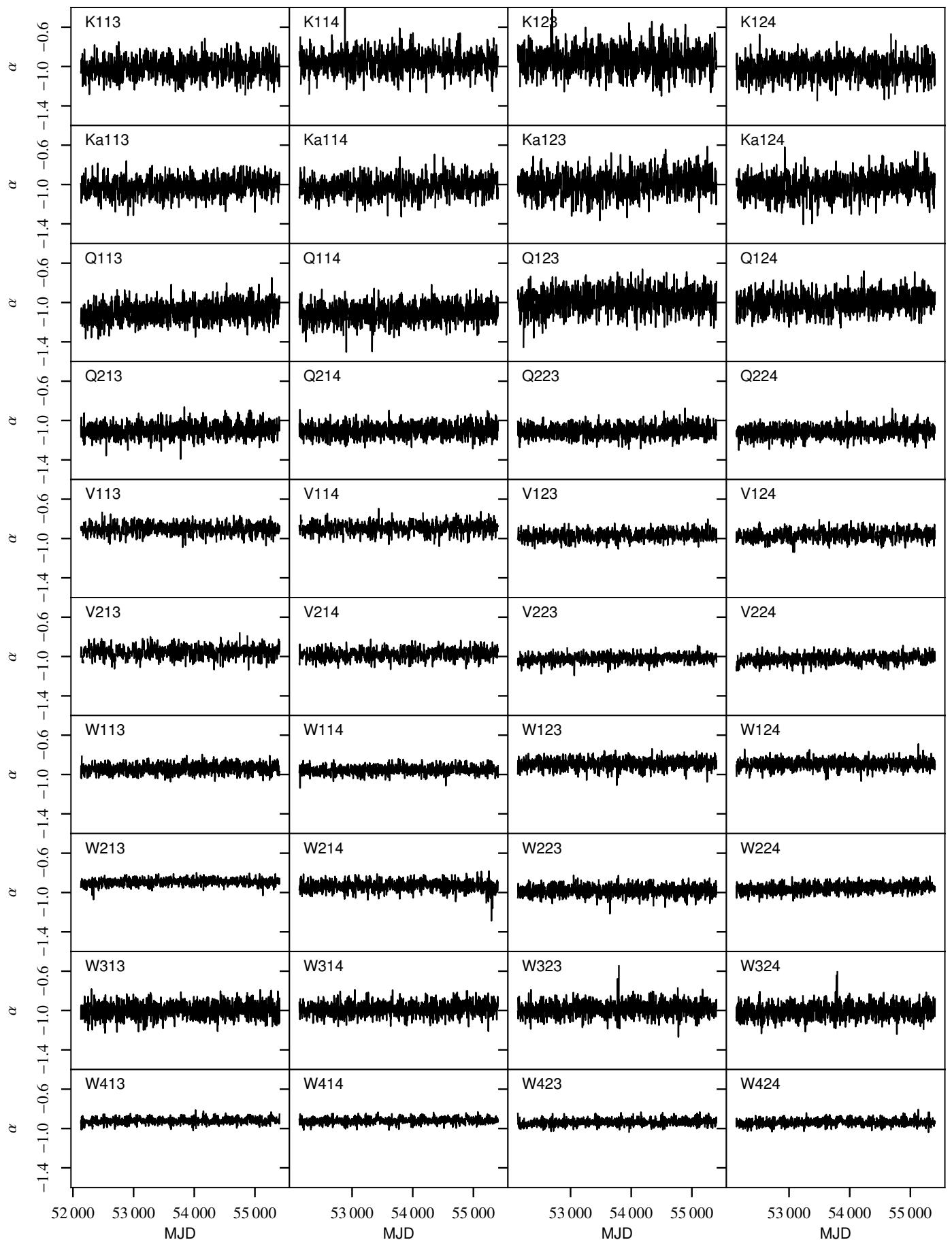
**Fig. A.2.** baseline slopes.

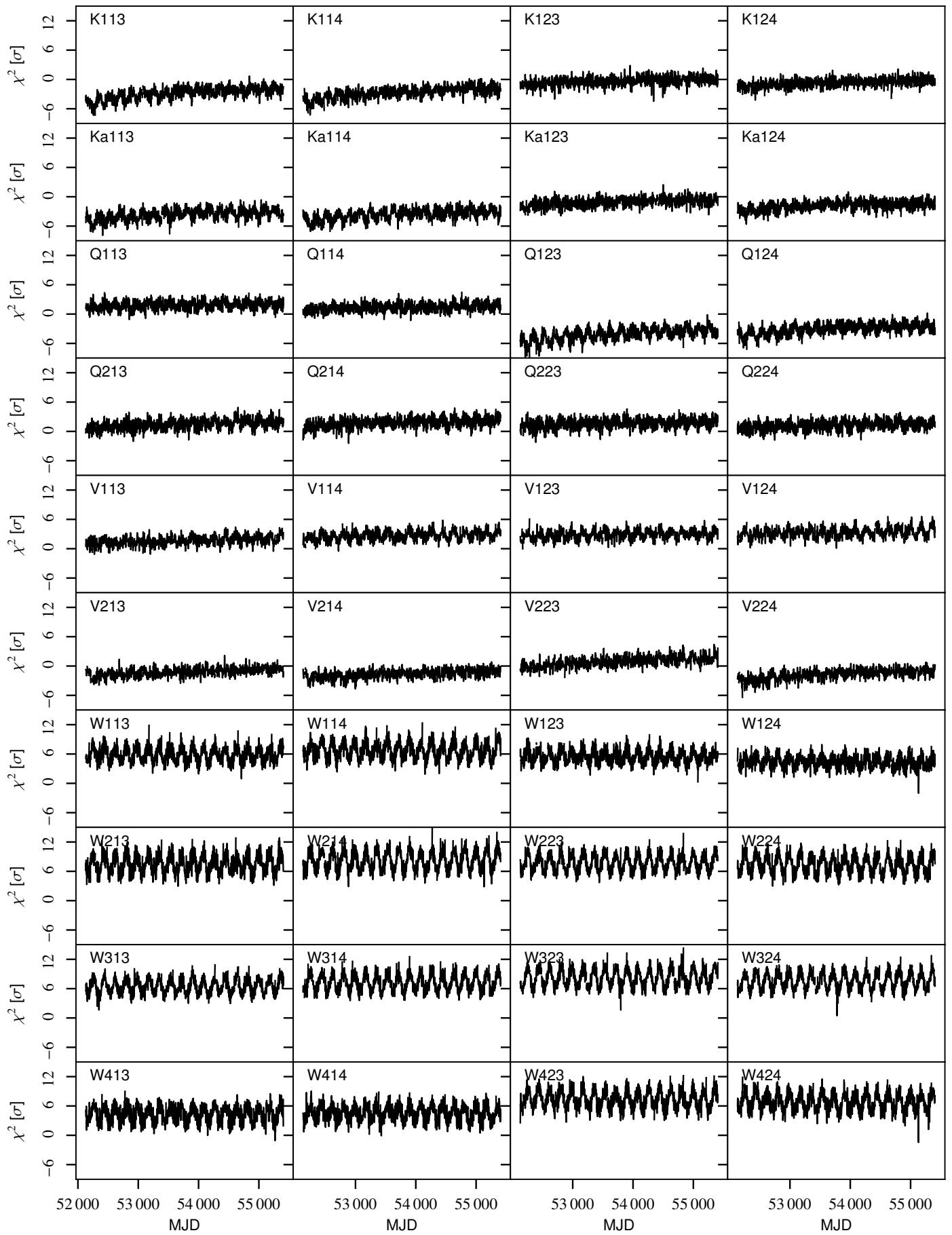
**Fig. A.3.** Gain.

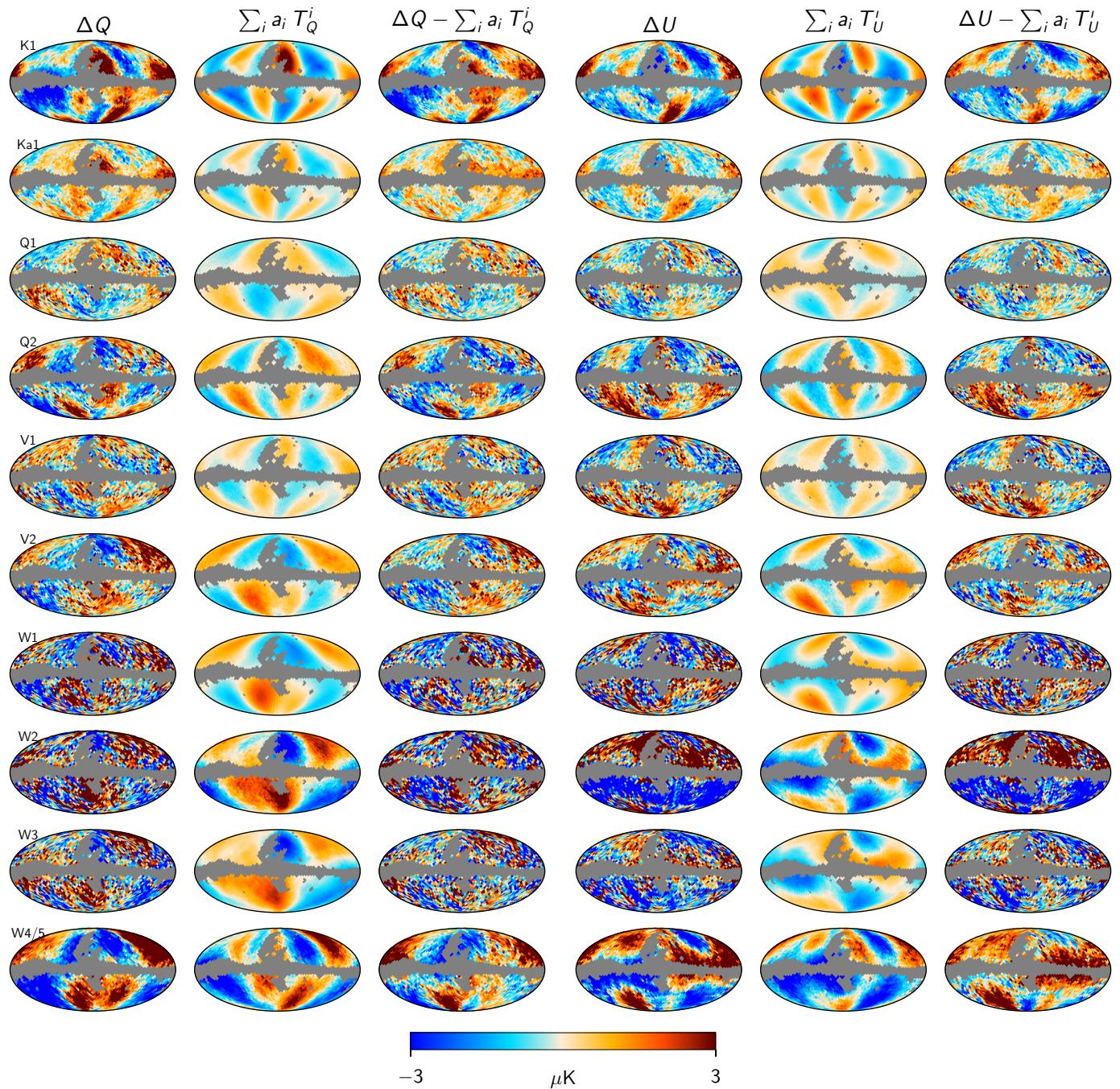
**Fig. A.4.** Relative difference in gain solutions,  $(g^{\text{CG}} - g^{\text{WMAP}})/g^{\text{WMAP}}$ .

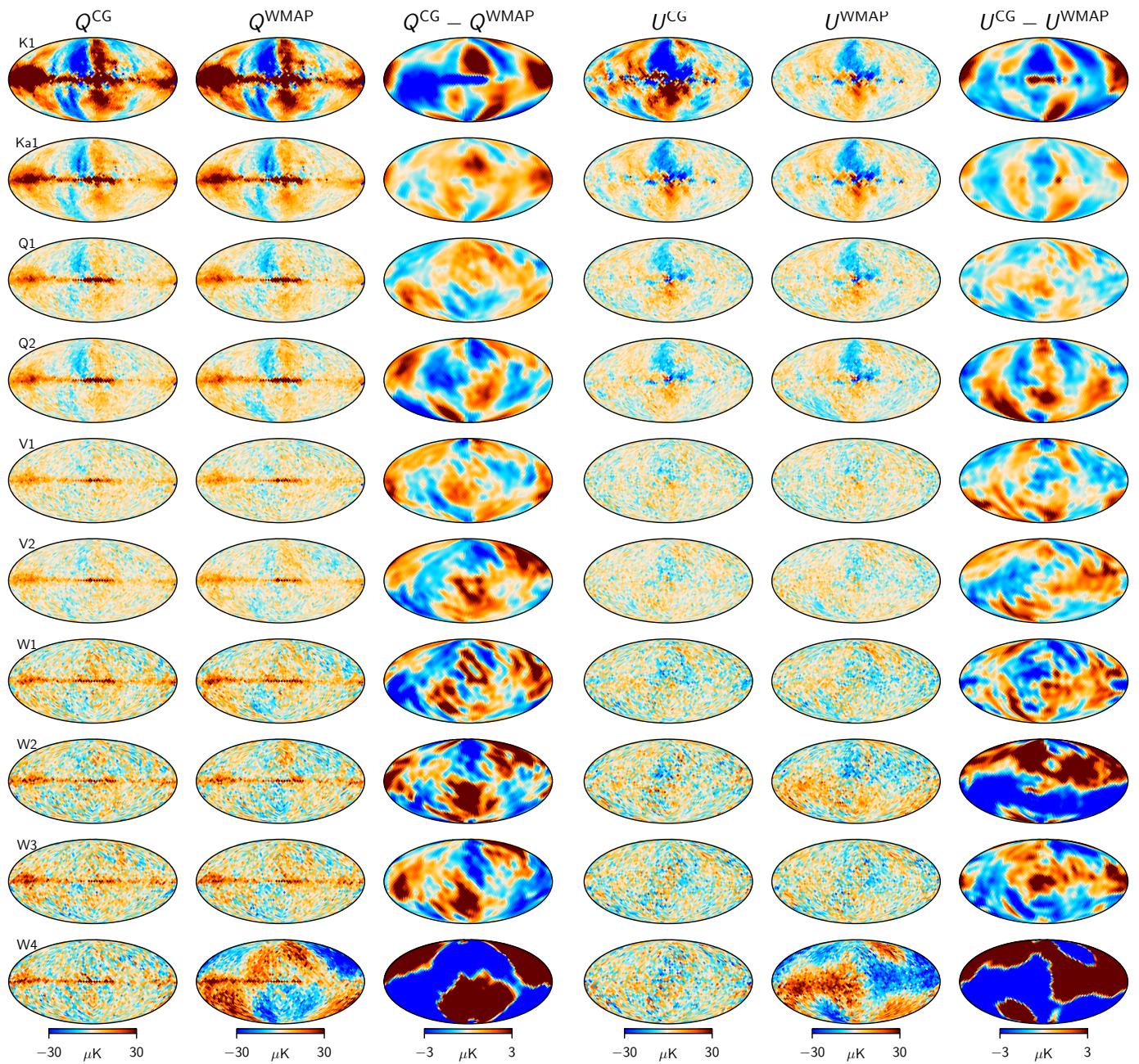
**Fig. A.5.**  $\sigma_0$ .

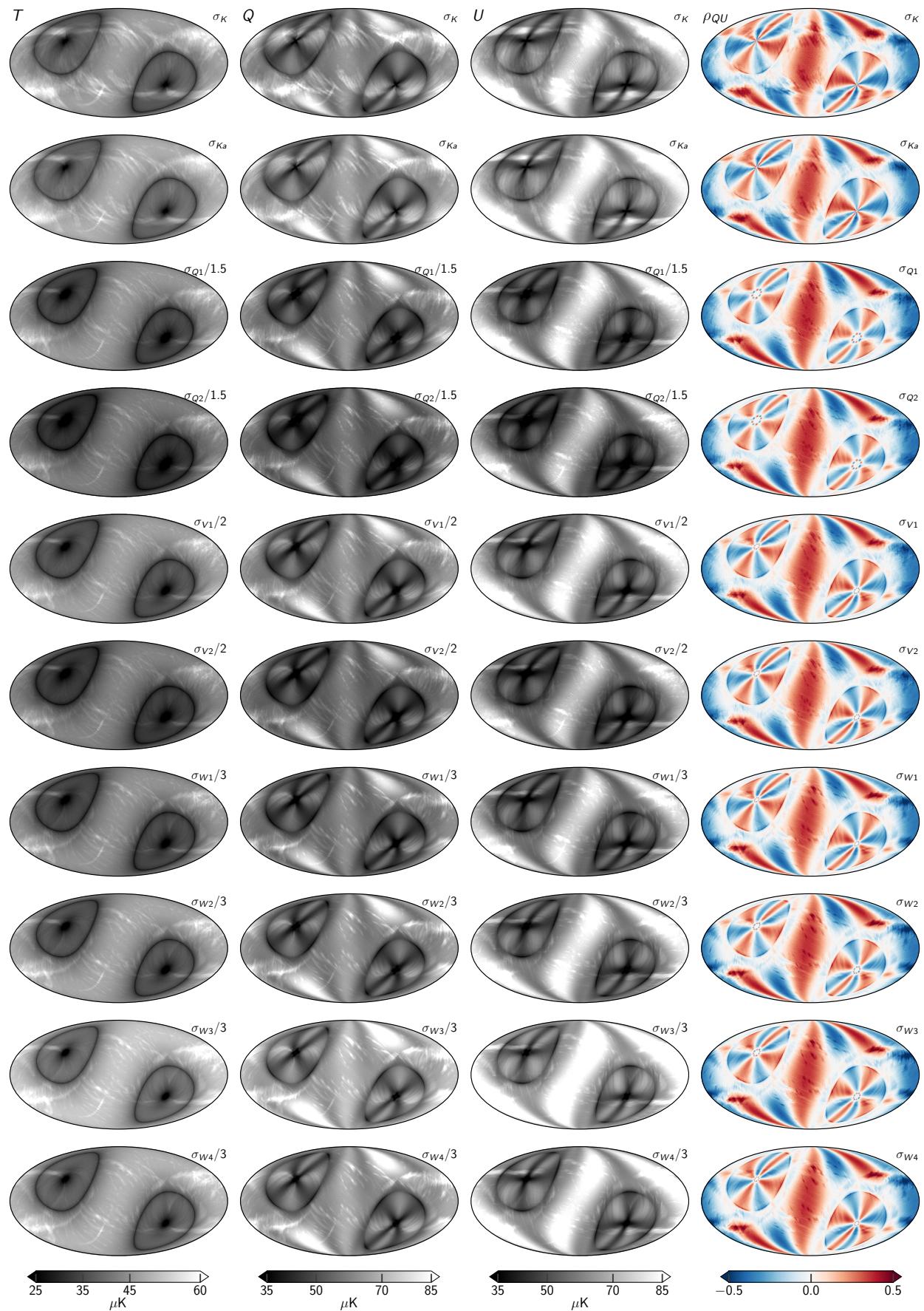
**Fig. A.6.** Fknee.

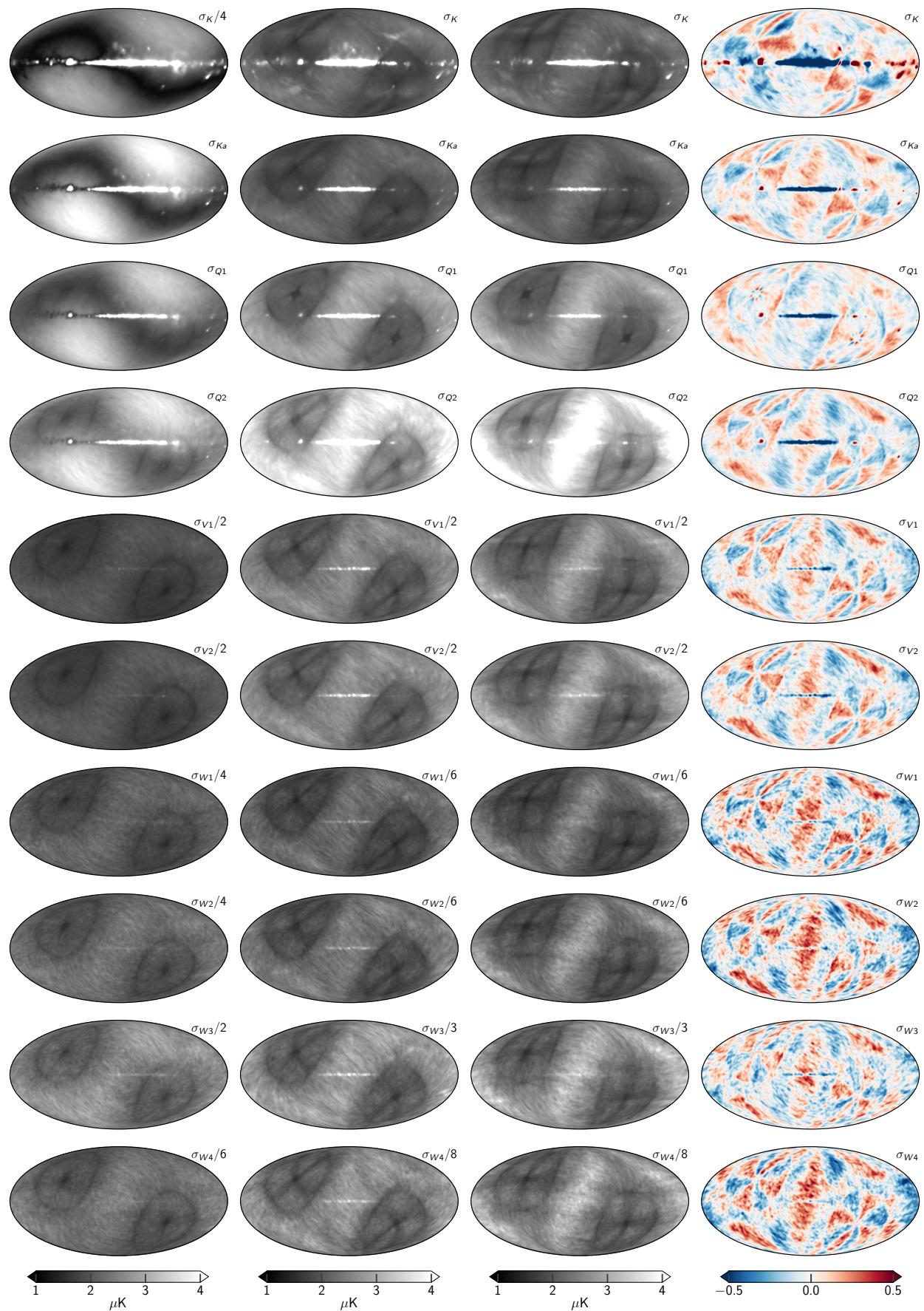
**Fig. A.7.** alpha.

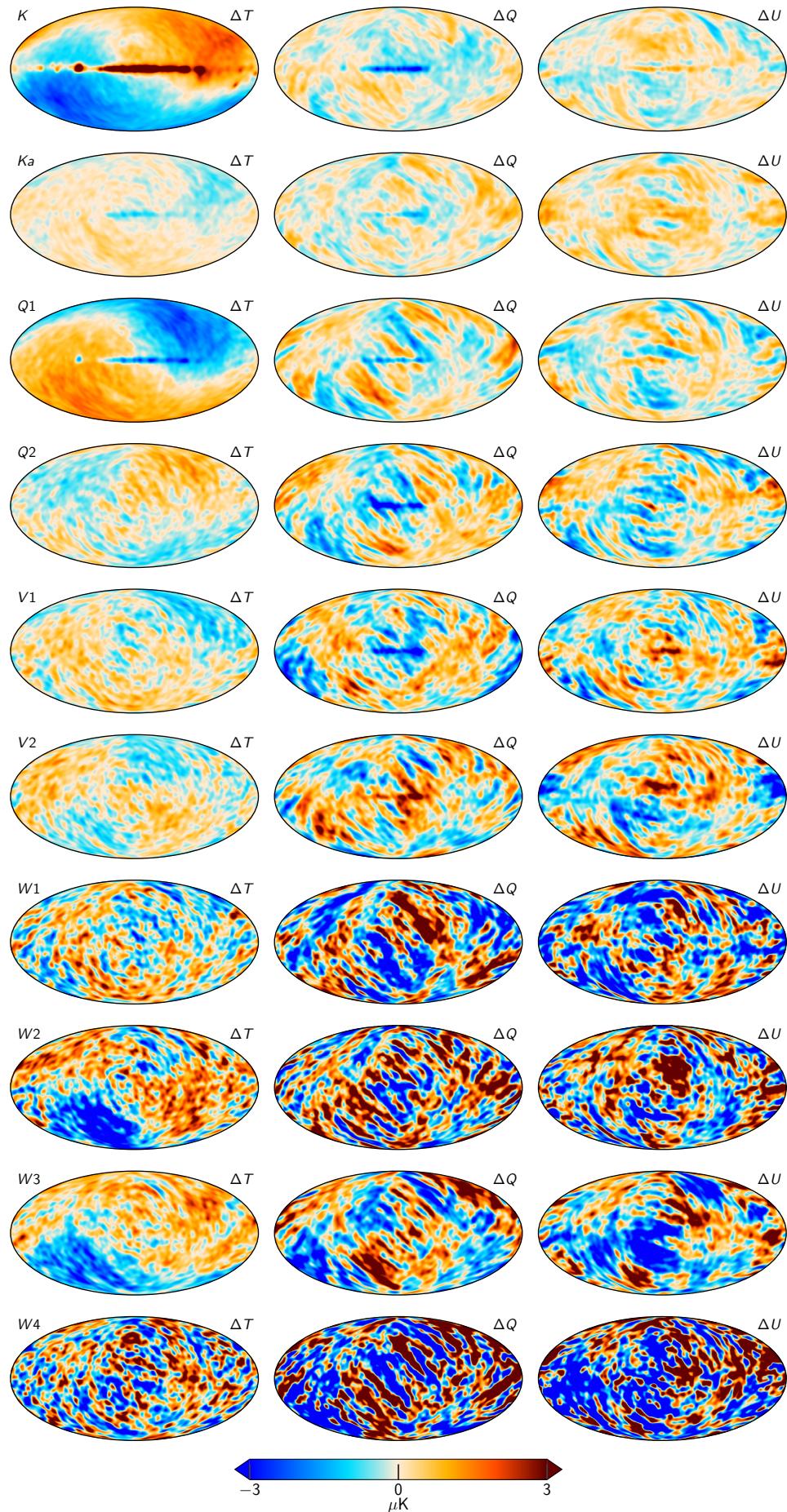
**Fig. A.8.** chisq.

**Fig. A.9.** Transmission imbalance templates

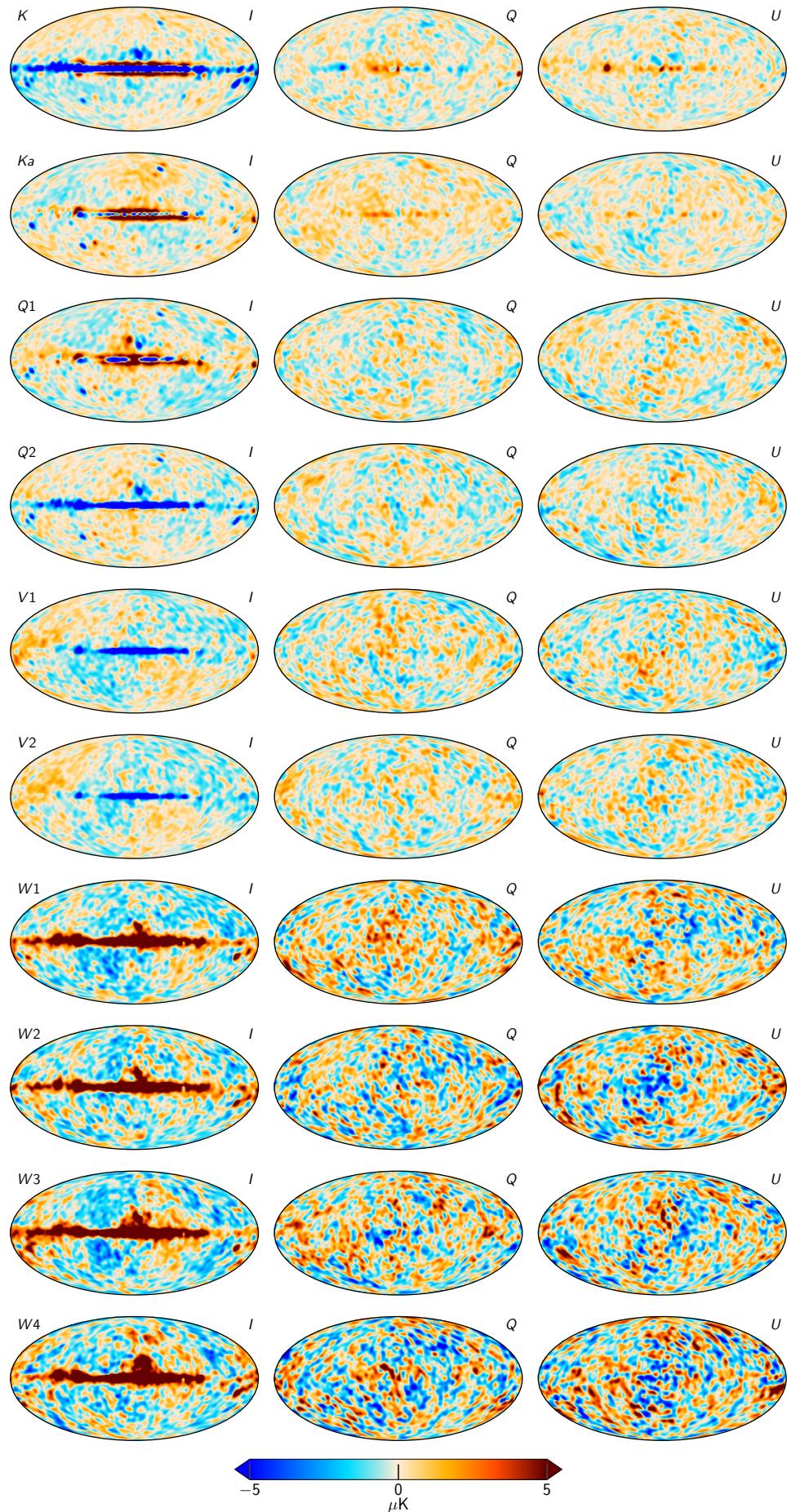
**Fig. B.1.** Sky maps

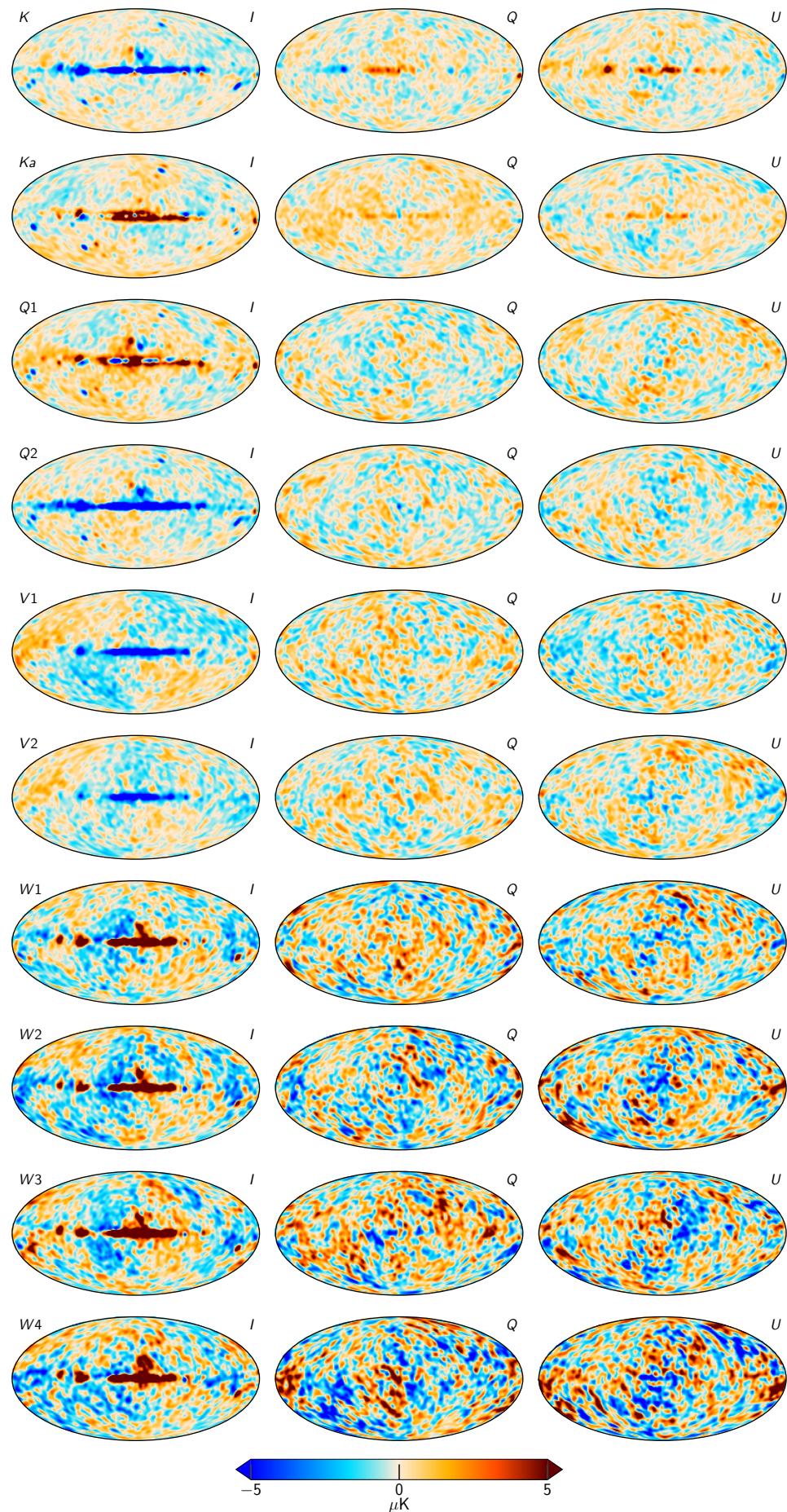
**Fig. B.2.** RMS maps

**Fig. B.3.** STD std



**Fig. B.4.** Differences between two samples  
Article number, page 56 of 58

**Fig. B.5.** TOD Residuals for each of the WMAP channels, smoothed by  $5^\circ$ .



**Fig. B.6.** Component separation residuals for each of the WMAP channels, smoothed by  $5^\circ$ .  
Article number, page 58 of 58