

COSMOGLOBE DR 1. I. Improved Wilkinson Microwave Anisotropy Probe frequency maps by Bayesian end-to-end analysis

D. J. Watts^{1*}, A. Basyrav¹, M. Galloway¹, D. Herman¹, H. T. Ihle¹, S. Paradiso², F. Rahman³, H. Thommesen¹, M. Bersanelli⁴, L. A. Bianchi⁴, M. Brilenkov¹, L. P. L. Colombo⁴, H. K. Eriksen¹, J. R. Eskilt^{1,5}, C. Franceschet⁴, U. Fuskeland¹, E. Gjerløw¹, B. Hensley⁷, L. T. Herzt⁸, G. A. Hoerning⁶, K. Lee¹, J. G. S. Lunde¹, A. Marins^{6,9}, S. K. Nerval^{10,11}, S. K. Patel¹², M. Regnier¹³, M. San¹, S. Sanyal¹², N.-O. Stutz¹, A. Verma¹², I. K. Wehus¹, and Y. Zhou¹⁴

¹ Institute of Theoretical Astrophysics, University of Oslo, Blindern, Oslo, Norway

² Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada

³ Indian Institute of Astrophysics, Koramangala II Block, Bangalore, 560034, India

⁴ Dipartimento di Fisica, Università degli Studi di Milano, Via Celoria, 16, Milano, Italy

⁵ Imperial Centre for Inference and Cosmology, Department of Physics, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, United Kingdom

⁶ Instituto de Física, Universidade de São Paulo - C.P. 66318, CEP: 05315-970, São Paulo, Brazil

⁷ Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08540

⁸ Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver BC, V6T1Z1, Canada

⁹ Department of Astronomy, University of Science and Technology of China, Hefei, China

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

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

¹² Department of Physics, Indian Institute of Technology (BHU), Varanasi - 221005, India

¹³ Laboratoire Astroparticule et Cosmologie (APC), Université Paris-Cité, Paris, France

¹⁴ Department of Physics, UC Berkeley

March 7, 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 581 CPU-hr, including calibration, mapmaking, and component separation, of which 389 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 $\sigma_2 = 120 \pm 65 \mu\text{K}^2$, 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

1 Introduction	2	2.4 COSMOGLOBE instrument model	6
2 End-to-end Bayesian CMB analysis	3	2.5 Sky model	6
2.1 LFI and BEYONDPLANCK	3	2.6 Priors and poorly measured modes	8
2.2 Official <i>WMAP</i> instrument model and analysis pipeline	3	2.7 Posterior distribution and Gibbs sampling	9
2.3 Instrument model	4	2.8 Sampling algorithms	9
		2.8.1 Review of sampling algorithms	9
		2.8.2 Differential mapmaking	10
		2.8.3 Transmission imbalance estimation	11
		2.8.4 Baseline sampling	11
3 Data and data processing			11

* Corresponding author: D. J. Watts; duncanwa@astro.uio.no

3 Data and data processing

3.1	Publicly available <i>WMAP</i> products	12
3.2	TOD pre-processing and data selection	12
3.3	Computational resources and future analysis plan	13
4	Instrumental parameters	13
4.1	Trace plots and correlations	13
4.2	Gain and baselines	13
4.3	Transmission imbalance	15
4.4	Instrumental noise and goodness-of-fit	16
5	Frequency maps	16
5.1	Map summary statistics	16
5.2	Comparison with 9-year <i>WMAP</i> maps	24
5.3	Consistency within <i>WMAP</i> channels	28
5.4	Consistency between <i>WMAP</i> and LFI	29
6	Preliminary astrophysical results	30
6.1	CMB results	30
6.1.1	Solar dipole	30
6.1.2	Angular temperature power spectrum . .	31
6.1.3	Low- ℓ anomalies	32
6.2	Galactic foregrounds	34
6.3	<i>WMAP</i> -versus-LFI signal-to-noise ratio comparison	35
7	Systematic error corrections and uncertainties	35
7.1	Sky map corrections	35
7.2	Power spectrum residuals	35
8	Outstanding issues	35
8.1	Noise modeling	35
8.2	<i>V</i> -band quadrupole residual	35
8.3	Degeneracy between <i>K</i> -band calibration and AME dipole	35
8.4	Other minor effects	35
8.4.1	Time-variable bandpass modeling	35
8.4.2	Polarized sidelobe modeling	35
9	Conclusions	35
A	Survey of instrumental parameters	39
A.1	Gain, baselines, noise and χ^2	39
A.2	Transmission imbalance	39
B	<i>WMAP</i> frequency map survey	39
C	Comparison with BEYONDPLANCK LFI results	39

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 Λ 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 COSMOGLOBE¹ 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, and this constitutes the first COSMOGLOBE Data Release (DR1). In doing so, this paper in fact 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 the global posterior distribution, are described in Sects. 4–6, summarizing instrumental parameters, frequency sky maps, and astrophysical results, respectively. In Sect. 7 we quantify the systematic error budget for this analysis, while we address a few minor unresolved issues in Sect. 8 that should be addressed in future work. We conclude in Sect. 9, 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 realtive gain uncertainty modes. Indeed, already the PR3 analysis 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, these 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 by *WMAP*.

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 DA is labelled, Xn , where X is the band label and n is index of the DA in question, e.g., *K1* or *W3*. A second digit m is used to denote whether which linear polarization, “1” or “2”, the radiometer is sensitive to. Additionally, each detector is a two-channel device that is internally left or right “jammed” with a 180° phase lag in the electronics (Jarosik et al. 2003b). These datastreams are further labeled by an integer k that takes values “3” or “4” (Hinshaw et al. 2003). For example, the label *Q223* refers to the second *Q*-band DA with corresponding to the second polarization orientation’s first channel.

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 in Bennett et al. (2013), there has been an observed drift in the center frequency of *K*, *Ka*, *Q*, and *V*-band corresponding to a ~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.²

The beams were characterized in the form of maps, with separate products for the near-field and far-field. The main beam

¹ <https://cosmoglobe.uio.no>

² https://lambda.gsfc.nasa.gov/product/wmap/dr5/m_sw.html

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). 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 housekeeping data, the quaternion pointing solution was corrected using a linear model that depends on arcsecond per kelvin. 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° downwards. In total, when horn A was pointed at pixel p_A , horn B was pointed at a pixel p_B approximately 141° 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

$$\begin{aligned} d_t^{\text{imbal}} &\propto (1 + x_{\text{im}})T_{p_A} - (1 - x_{\text{im}})T_{p_B} \\ &= (T_{p_A} - T_{p_B}) + x_{\text{im}}(T_{p_A} + T_{p_B}), \end{aligned} \quad (1) \quad (2)$$

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_{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 \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

$$\begin{aligned} \mathbf{d}_2 &= \mathbf{P}_2 \mathbf{m} \\ &= (1 + x_{\text{im}})[T_A - Q_A \cos 2\gamma_A - U_A \sin 2\gamma_A - S_A] \\ &\quad + (1 - x_{\text{im}})[-T_B + Q_B \cos 2\gamma_B + U_B \sin 2\gamma_B + S_B]. \end{aligned} \quad (5) \quad (6)$$

In this formalism, S acts as an extra Stokes parameter that absorbs the effects of differing bandpass responses between radiometers \mathbf{d}_1 and \mathbf{d}_2 (Jarosik et al. 2007). The asymmetric mapmaking matrix, \mathbf{P}_{am} , was used because, as noted in 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° 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 $N_{tt'}^{-1}$. 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. Instrument model

As opposed to the WMAP pipeline, the COSMOGLOBE approach uses a generative model for every aspect of the data, including the sky and instrumental effects at once. In the COSMOGLOBE paradigm, it is simplest to characterize the data's goodness-of-fit to a model,

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

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}^{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 $f_c(\nu/\nu_{0,c})$ and a detector j with bandpass $\tau_j(\nu)$, given by

$$\mathbf{M}_{cj} = \int d\nu \tau_j(\nu) f_c(\nu/\nu_{c,0}). \quad (8)$$

The maps \mathbf{a} are the component amplitudes, \mathbf{s}^{orb} is the orbital dipole induced by the motion of the telescope with respect to the Sun, and \mathbf{s}^{fsi} is the time-dependent far sidelobe signal. In the Commander3 (Galloway et al. 2022a) implementation, \mathbf{n}^{corr} is a realization of the correlated noise component whose power spectrum is parametrized by $P(f \mid \xi_n)$, where ξ_n generally includes the knee frequency f_{knee} , a slope α , and an amplitude fixed by the white noise σ_0 . Similarly, each experiment has particular signals that are specific to the instrument in question, e.g., the analog-to-digital converter nonlinearities (Herman et al. 2022a) and 1 Hz spike in *Planck* LFI or the large baseline in *WMAP*, which can be modeled by \mathbf{s}^{inst} .

The bandpass mismatch is explicitly modeled in Commander3. Using the calibrated sky model, the expected calibrated sky signal 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 maps \mathbf{m}_j will each be slightly different depending on their bandpass τ_j . More importantly, differences in signal between different detectors can be attributed to a spurious polarization signal, giving spurious polarized maps. Therefore, before averaging different detectors together, Commander3 estimates 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.

The most unique feature of the *WMAP* data is that it includes differential pointing. *WMAP* has two primary mirrors approximately reflected along the vertical satellite axis, and are tilted approximately 19.5° downwards – in total, when horn A is pointed at pixel p_A , horn B is pointed at a pixel p_B that is approximately 141° away (Page et al. 2003). The incoming radiation is 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 has a partner pair of radiometers that observes 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

$$d_t = g_t \mathbf{P}_{t,p} s_p + s_t^{\text{inst}} + n_t \quad (11)$$

$$\begin{aligned} &= g_t [\alpha_A (T_{p_A(t)} + Q_{p_A(t)} \cos 2\gamma_A(t) + U_{p_A(t)} \sin 2\gamma_A(t)) \\ &\quad - \alpha_B (T_{p_B(t)} + Q_{p_B(t)} \cos 2\gamma_B(t) + U_{p_B(t)} \sin 2\gamma_B(t))] \\ &\quad + n_t, \end{aligned} \quad (12)$$

where s_p is the total sky signal $\mathbf{B}\mathbf{m}$, $\gamma_{A/B}$ are the time-dependent polarization angles, and $p_{A/B}$ are the A and B pixel locations.

In this notation, $\alpha_{A/B}$ is the total optical power transmission for horn A and B.

For an ideal radiometer, $\alpha_B = \alpha_B$. However, for a real detectors the two transmissions vary slightly, and this effect is called “transmission imbalance”. Defining the transmission imbalance for a given radiometer pair as $x_{\text{im}} \equiv (\alpha_A - \alpha_B)/(\alpha_A + \alpha_B)$, and absorbing $\alpha_A + \alpha_B$ into the definition of g_t , allows the pointing operation to be written in terms of the deviation from ideality,

$$\begin{aligned} \mathbf{P}_{t,p} s_p &= (1 + x_{\text{im}})(T_{p_A} + Q_{p_A} \cos 2\gamma_A + U_{p_A} \sin 2\gamma_A) \\ &\quad - (1 - x_{\text{im}})(T_{p_B} + Q_{p_B} \cos 2\gamma_B + U_{p_B} \sin 2\gamma_B). \end{aligned} \quad (13)$$

It is important to note that inaccuracies in the determination of x_{im} will yield a spurious polarization component, effectively leaking the bright temperature signal into the polarization maps. In particular coupling with the bright Solar CMB dipole is an important effect that give rise to polarization-specific artefacts with a clear signature defined by relative orientation of the Solar dipole and the *WMAP* scanning strategy (Jarosik et al. 2007). To account for this effect, the *WMAP* pipeline fit x_{im} to the Solar dipole in TOD space, while simultaneously accounting for both common and differential TOD 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, observations near planets 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} , calibrated data of the form $\mathbf{d} = (\mathbf{d}_{\text{raw}} - \mathbf{b})/\mathbf{g}$ were fed into the asymmetric mapmaking equation,

$$\mathbf{P}_{\text{am}}^T \mathbf{N}^{-1} \mathbf{P} \mathbf{m} = \mathbf{P}_{\text{am}}^T \mathbf{N}^{-1} \mathbf{d}, \quad (14)$$

where \mathbf{N} is the time-domain covariance matrix and \mathbf{P} is the pointing matrix implicitly defined for detectors \mathbf{d}_1 and \mathbf{d}_2 sensitive to different polarization orientations as follows,

$$\begin{aligned} \mathbf{d}_1 &= \mathbf{P}_1 \mathbf{m} \\ &= (1 + x_{\text{im}})[T_A + Q_A \cos 2\gamma_A + U_A \sin 2\gamma_A + S_A] \\ &\quad + (1 - x_{\text{im}})[-T_B - Q_B \cos 2\gamma_B - U_B \sin 2\gamma_B - S_B], \end{aligned} \quad (15)$$

and

$$\begin{aligned} \mathbf{d}_2 &= \mathbf{P}_2 \mathbf{m} \\ &= (1 + x_{\text{im}})[T_A - Q_A \cos 2\gamma_A - U_A \sin 2\gamma_A - S_A] \\ &\quad + (1 - x_{\text{im}})[-T_B + Q_B \cos 2\gamma_B + U_B \sin 2\gamma_B + S_B]. \end{aligned} \quad (16)$$

In this formalism, S acts as an extra Stokes parameter that absorbs the effects of differing bandpass responses between radiometers \mathbf{d}_1 and \mathbf{d}_2 (Jarosik et al. 2007). The asymmetric mapmaking matrix, \mathbf{P}_{am} , was used because, as noted in Jarosik et al. (2011), large gradients within a single pixel observed in one beam, for instance due to bright Galactic foregrounds or point sources, could leak into the solution for the pixel observed by the other beam, leading to artefacts 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. The noise covariance matrix, \mathbf{N} , was modelled in terms of the autocorrelation function estimated separately for each year of data. Finally, to create the sky maps themselves, the data were processed one year at a time, applying the iterative Bi-conjugate Gradient Stabilized Method (BiCG-STAB van der Vorst 1992; Barrett et al. 1994) to solve for the actual maps.

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

$$\mathbf{d} = \mathbf{GP}[\mathbf{B}^{\text{symm}}\mathbf{Ma} + \mathbf{B}^{4\pi}(s^{\text{orb}} + s^{\text{fsl}})] + s^{\text{inst}} + \mathbf{n}^{\text{corr}} + \mathbf{n}^{\text{w}}, \quad (17)$$

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}^{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 $f_c(\nu/\nu_{0,c})$ and a detector j with bandpass $\tau_j(\nu)$, given by

$$\mathbf{M}_{cj} = \int d\nu \tau_j(\nu) f_c(\nu/\nu_{c,0}). \quad (18)$$

The maps \mathbf{a} represent the Stokes parameters for each astrophysical component, while s^{orb} is the orbital dipole induced by the motion of the telescope with respect to the Sun, and s^{fsl} is the time-dependent far sidelobe signal. Following Ihle et al. (2022), we model the correlated noise component \mathbf{n}^{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 σ_0 denotes the white noise amplitude, f_{knee} is the so-called $1/f$ knee frequency, and α 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 χ^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. 8.1.

The term s^{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 $s_t^{\text{WMAP}} = b_0 + b_1 t$, 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 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}^{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

$$m_{p,j} = \mathbf{B}_{p,p'} \sum_c \mathbf{M}_{c,j} a_{p'}^c + n_{j,p}^{\text{w}}. \quad (19)$$

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 τ_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;

$$\delta s_{t,j}^{\text{leak}} = \mathbf{P}_{t,p}^j \mathbf{B}_{p,p'}^j (\mathbf{m}_{j,p'} - \mathbf{m}_{p'}). \quad (20)$$

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}_{raw} , is displayed in the top panel, with the sky signal $\mathbf{s}_{\text{sky}} = \mathbf{PB}^{\text{symm}}\mathbf{Ma}$ plotted directly underneath. The next four panels show the correlated noise realization \mathbf{n}_{corr} , the orbital dipole \mathbf{s}_{orb} , the far sidelobe contribution \mathbf{s}_{sl} , and the bandpass leakage \mathbf{s}_{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.

2.5. Sky model

Following BeyondPlanck (2022), we assume that the sky (as defined by \mathbf{Ma} in Eq. 17) across the *WMAP* frequencies can be modeled as a linear combination of CMB fluctuations (\mathbf{a}_{CMB}), synchrotron (\mathbf{a}_{s}), free-free emission (\mathbf{a}_{ff}), anomalous microwave emission (\mathbf{a}_{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,

$$\mathbf{s}_{\text{RJ}} = (\mathbf{a}_{\text{CMB}} + \mathbf{a}_{\text{quad}}(\nu)) \frac{x^2 e^x}{(e^x - 1)^2} + \quad (21)$$

$$+ \mathbf{a}_{\text{s}} \left(\frac{\nu}{\nu_{0,\text{s}}} \right)^{\beta_{\text{s}}} + \quad (22)$$

$$+ \mathbf{a}_{\text{ff}} \left(\frac{\nu_{0,\text{ff}}}{\nu} \right)^2 \frac{g_{\text{ff}}(\nu; T_{\text{e}})}{g_{\text{ff}}(\nu_{0,\text{ff}}; T_{\text{e}})} + \quad (23)$$

$$+ \mathbf{a}_{\text{ame}} e^{\beta_{\text{AME}}(\nu - \nu_{0,\text{ame}})} + \quad (24)$$

$$+ \mathbf{a}_{\text{d}} \left(\frac{\nu}{\nu_{0,\text{d}}} \right)^{\beta_{\text{d}}+1} \frac{e^{h\nu_{0,\text{d}}/k_{\text{B}}T_{\text{d}}} - 1}{e^{h\nu/k_{\text{B}}T_{\text{d}}} - 1} + \quad (25)$$

$$+ U_{\text{mJy}} \sum_{j=1}^{N_{\text{src}}} \mathbf{a}_{j,\text{src}} \left(\frac{\nu}{\nu_{0,\text{src}}} \right)^{\alpha_{j,\text{src}}-2}, \quad (26)$$

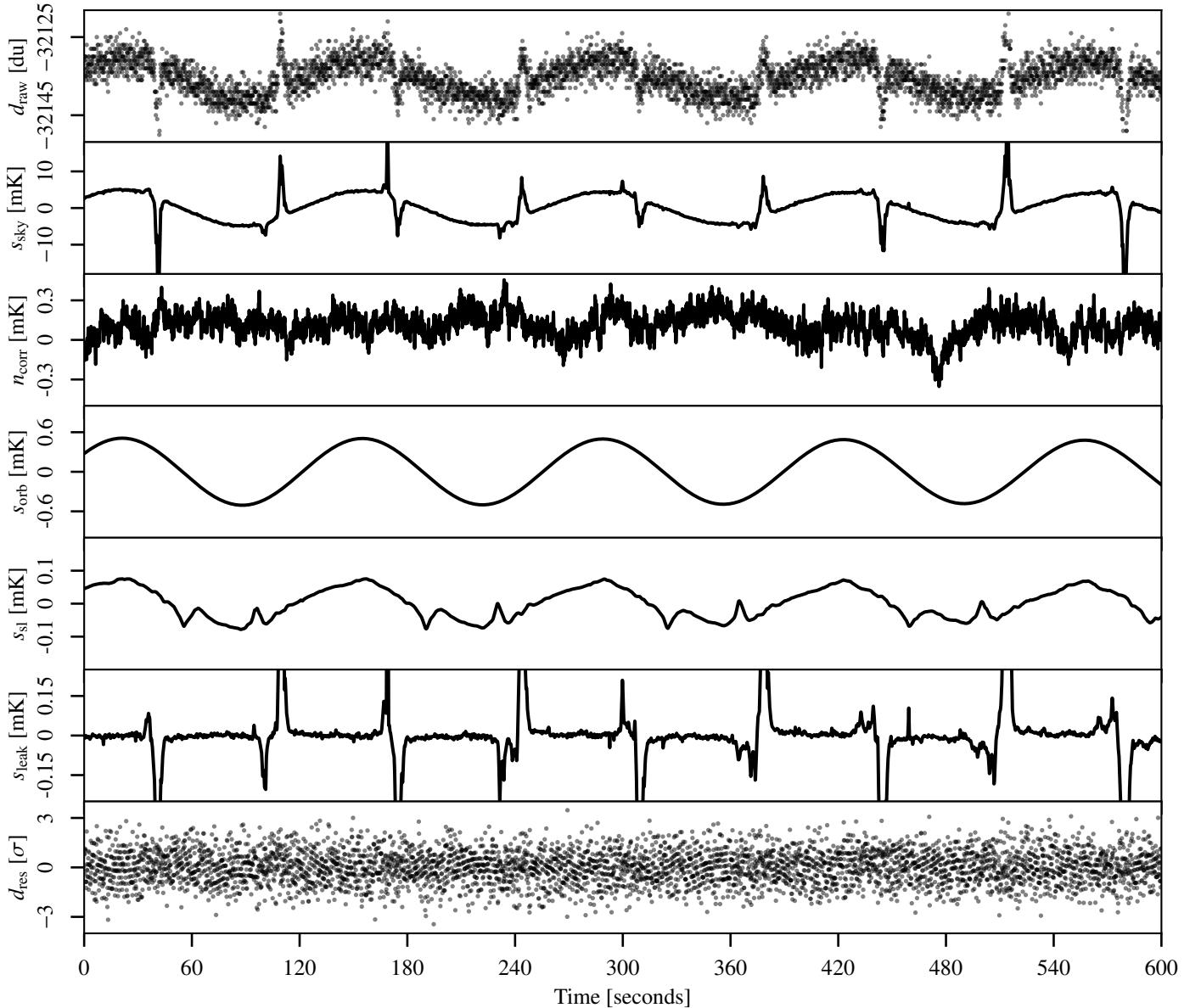


Fig. 1. Time-ordered data segment for the K113 radiometer. From top to bottom, the panels show 1) raw uncalibrated TOD \mathbf{d} ; 2) sky signal s_{sky} ; 3) calibrated correlated noise n_{corr} ; 4) orbital CMB dipole signal s_{orb} ; 5) sidelobe correction s_{sl} ; 6) bandpass leakage correction s_{leak} ; 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.

where $x = h\nu/kT_{\text{CMB}}$; $\nu_{0,i}$ is a reference frequency for component i ; β_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_{ff} is the so-called Gaunt factor [Dickinson et al. \(2003\)](#); β_{AME} is an exponential scale factor for AME emission (see below); β_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_{mJy} is the conversion factor between flux density (in millijansky) and brightness temperature (in K_{RJ}) for the channel in question. Also, a_{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-

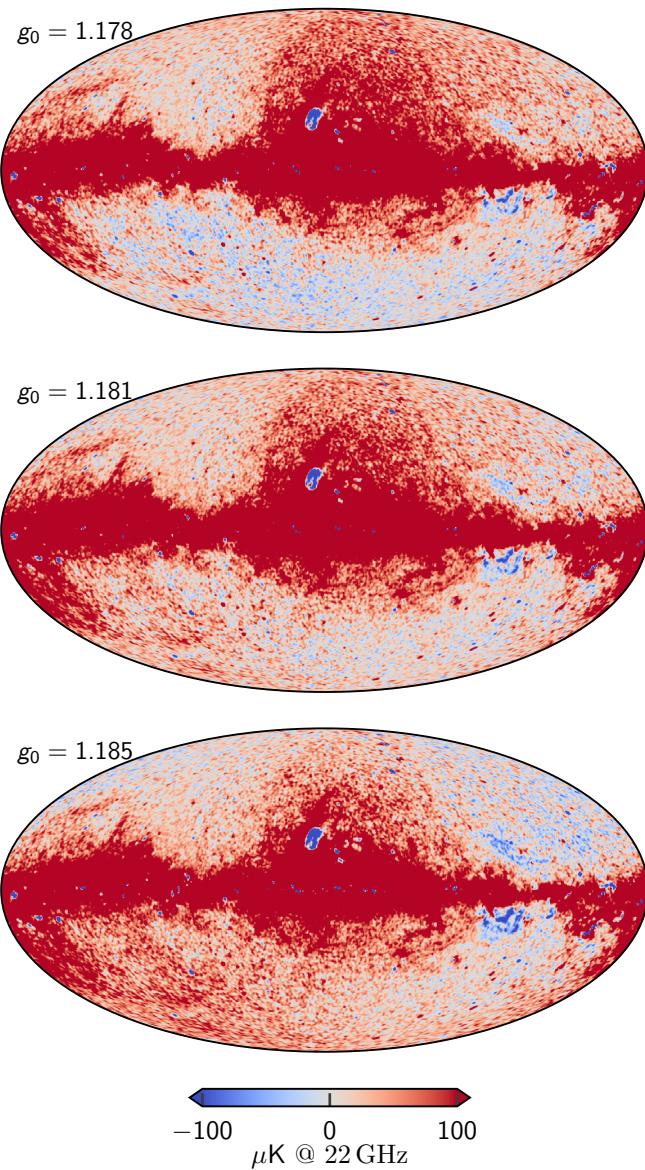


Fig. 2. Dependence on AME amplitude evaluated at 22 GHz as a function of absolute calibration. Each map comes from the fifth iteration of a dedicated `Commander` run that fixed g_0 while letting all other TOD parameters be fit. The values of $g_0 = 1.178$ and $g_0 = 1.185$ represent 3.5σ draws from the 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.

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.

2.6. Priors and poorly measured modes

The model described in Sects. 2.4 and 2.5 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-

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 \(2018\)](#); [Paradiso et al. \(2022\)](#), which dictates that astrophysical foregrounds have 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, β_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. 8.3, 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 intuitively 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 χ^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 $\mu\text{K} \text{ mK}^{-1}$; the extreme values differ only by 0.25 %. All of these three values appear equally acceptable from a pure χ^2 point-of-view, relative to the noise level and modelling errors of these data. At the same time, it is obviously 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. It is then important to recall 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 obvious advantage of such a single-channel approach is that the solution is independent of the assumed sky model. However, an equally obvious disadvantage is that it is impossible to break any potential inherent degeneracies with that approach; it cannot be combined with external observations in any meaningful way. And one important degeneracy 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 $N(-3.30, 0.1^2)$, 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 $\beta_{\text{AME}} = N(3.56, 0.1)$. The parameters of these priors were determined by running a grid over β_{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 χ^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 all-important 11–19 GHz frequency range, suggests that the entire AME model should be revisited in a future joint *WMAP*+LFI+QUIJOTE analysis, and 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).

2.7. 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, $\mathbf{d} = \mathbf{s}^{\text{tot}}(\omega) + \mathbf{n}^w$, where \mathbf{s}^{tot} encompasses all of the terms in Eq. (17) 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 σ_0^2 . In general, the likelihood reads

$$P(\mathbf{d} | \omega) \propto \exp\left(-\frac{1}{2} \sum_t \frac{(d_t - s_t^{\text{tot}}(\omega))^2}{\sigma_0^2}\right). \quad (27)$$

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 χ^2 -distribution with n_{TOD} degrees of freedom. In the limit of large n , a χ^2 distribution is well-approximated by a Gaussian with mean n and variance $2n$. Therefore we define and use in the following the reduced normalized χ^2 statistic,

$$\chi^2 \equiv \frac{\sum_t ((d_t - s_t^{\text{tot}})^2 / \sigma_0^2) - n_{\text{TOD}}}{\sqrt{2n_{\text{TOD}}}}, \quad (28)$$

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

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

$$\mathbf{g} \leftarrow P(\mathbf{g} \mid \mathbf{d}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, C_\ell, \theta) \quad (29)$$

$$\mathbf{n}_{\text{corr}} \leftarrow P(\mathbf{n}_{\text{corr}} \mid \mathbf{d}, \mathbf{g}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, C_\ell, \theta) \quad (30)$$

$$\xi_n \leftarrow P(\xi_n \mid \mathbf{d}, \mathbf{g}, \mathbf{n}_{\text{corr}}, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, C_\ell, \theta) \quad (31)$$

$$\mathbf{s}^{\text{inst}} \leftarrow P(\mathbf{s}^{\text{inst}} \mid \mathbf{d}, \mathbf{g}, \mathbf{n}_{\text{corr}}, \xi_n, \boldsymbol{\beta}, \mathbf{a}, C_\ell, \theta) \quad (32)$$

$$\boldsymbol{\beta} \leftarrow P(\boldsymbol{\beta} \mid \mathbf{d}, \mathbf{g}, \mathbf{n}_{\text{corr}}, \xi_n, \mathbf{s}^{\text{inst}}, C_\ell, \theta) \quad (33)$$

$$\mathbf{a} \leftarrow P(\mathbf{a} \mid \mathbf{d}, \mathbf{g}, \mathbf{n}_{\text{corr}}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, C_\ell, \theta) \quad (34)$$

$$C_\ell \leftarrow P(C_\ell \mid \mathbf{d}, \mathbf{g}, \mathbf{n}_{\text{corr}}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, \theta) \quad (35)$$

$$\theta \leftarrow P(\theta \mid \mathbf{d}, \mathbf{g}, \mathbf{n}_{\text{corr}}, \xi_n, \mathbf{s}^{\text{inst}}, \boldsymbol{\beta}, \mathbf{a}, C_\ell), \quad (36)$$

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.

2.8. 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. 2.8.1 we review the sampling algorithms implemented in the BEYONDPLANCK suite of papers, while Sects. 2.8.2–2.8.4 provide an overview of the *WMAP*-specific processing steps.

2.8.1. 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

$$d_{t,i} = g_{q,i} s_{t,i}^{\text{tot}} + n_{t,i}^{\text{corr}} + n_{t,i}^{\text{wn}} \quad (37)$$

may be almost equally well fit by two solutions defined by $g' = gs^{\text{tot}}/(s^{\text{tot}})'$ or $(n^{\text{corr}})' = n^{\text{corr}} + gs^{\text{tot}} + g'(s^{\text{tot}})'$; 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 \mathbf{n}^{corr} , and then drawing \mathbf{n}^{corr} from its conditional distribution with respect to g ,

$$\mathbf{g} \leftarrow P(\mathbf{g} \mid \mathbf{d}, \xi_n, \dots) \quad (38)$$

$$\mathbf{n}_{\text{corr}} \leftarrow P(\mathbf{n}_{\text{corr}} \mid \mathbf{d}, \mathbf{g}, \xi_n, \dots). \quad (39)$$

It is easy to see that this is a valid sample from the joint distribution simply from the definition of a conditional distribution, $P(\mathbf{g}, \mathbf{n}_{\text{corr}} \mid \omega) = P(\mathbf{n}_{\text{corr}} \mid \mathbf{g}, \omega)P(\mathbf{g} \mid \omega)$. 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}_{wn} .

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

$$g_{t,i} = g_0 + \Delta g_i + \delta g_{q,i} \quad (40)$$

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,

$$\mathbf{d}_i = \mathbf{g}_i s_i^{\text{tot}} + \mathbf{n}_i^{\text{tot}} \sim \mathcal{N}(\mathbf{g}_i s_i^{\text{tot}}, \mathbf{N}_i). \quad (41)$$

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 solving³

$$[(s_i^{\text{tot}})^T \mathbf{N}_i^{-1} s_i^{\text{tot}}] \mathbf{g}_i = (s_i^{\text{tot}})^T \mathbf{N}_i^{-1} \mathbf{d}_i + (s_i^{\text{tot}})^T \mathbf{N}_i^{-1/2} \boldsymbol{\eta}. \quad (42)$$

Note that the \mathbf{N}_i depends implicitly on the noise PSD ξ_n , while the specific realization of \mathbf{n}_{corr} is accounted for in the covariance matrix. As detailed by [Gjerløw et al. \(2022\)](#), **Commander3** samples in practice g_0 , Δg_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, Δg_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,

$$\mathbf{d}_i = \mathbf{g}_i s_i^{\text{tot}} + \mathbf{n}_i^{\text{corr}} + \mathbf{n}_i^{\text{wn}} \sim \mathcal{N}(\mathbf{g}_i s_i^{\text{tot}}, \mathbf{N}_{\text{corr},i} + \mathbf{N}_{\text{wn},i}). \quad (43)$$

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

$$(\mathbf{N}_{\text{corr},i}^{-1} + \mathbf{N}_{\text{wn},i}^{-1}) \mathbf{n}_i^{\text{corr}} = \mathbf{N}_{\text{wn},i}^{-1} \mathbf{r}_i + \mathbf{N}_{\text{wn},i}^{-1/2} \boldsymbol{\eta}_1 + \mathbf{N}_{\text{corr},i}^{-1/2} \boldsymbol{\eta}_2. \quad (44)$$

This gives a sample of the underlying correlated noise.

³ See, e.g., Appendix A.2 of [BeyondPlanck \(2022\)](#) for a derivation of this result.

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

$$P(\xi_n \mid \mathbf{n}^{\text{corr}}) \propto \frac{\exp[-\frac{1}{2}(\mathbf{n}^{\text{corr}})^T \mathbf{N}_{\text{corr}}^{-1} \mathbf{n}^{\text{corr}}]}{\sqrt{|\mathbf{N}_{\text{corr}}|}} P(\xi_n). \quad (45)$$

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

$$\mathbf{N}_{\text{corr}}(f) = \sigma_0^2 \left(\frac{f}{f_{\text{knee}}} \right)^\alpha. \quad (46)$$

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 σ_0 , we effectively fix the white noise level to the noise level at the highest frequency, e.g.,

$$\sigma_0^2 \equiv \frac{\text{Var}(r_{t+1} - r_t)}{2}, \quad (47)$$

where t and $t + 1$ are consecutive time samples, and $\mathbf{r} \equiv \mathbf{d} - \mathbf{g}s^{\text{tot}} - \mathbf{n}^{\text{corr}}$. In practice, this makes σ_0 a deterministic function of the sampled sky and gain parameters. The parameters α and f_{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,

$$r_{t,j} = \frac{d_{t,j} - n_{t,j}^{\text{corr}}}{g_{t,j}} - (s_{t,j}^{\text{orb}} + s_{t,j}^{\text{fsl}} + \delta s_{t,j}^{\text{leak}} + s_{t,j}^{\text{inst}}) \quad (48)$$

where s^{orb} is the orbital dipole ([Gjerløw et al. 2022](#)), s^{fsl} is the far sidelobe timestream ([Galloway et al. 2022b](#)), δs^{leak} is the bandpass leakage ([Svalheim et al. 2022a](#)), and s^{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.

2.8.2. 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

$$\mathbf{d} = \mathbf{P}\mathbf{m} + \mathbf{n}^{\text{w}}, \quad (49)$$

where

$$\mathbf{m} = \mathbf{B}^{\text{symm}} \mathbf{M} \mathbf{a} \quad (50)$$

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

$$\begin{aligned} \mathbf{P}_{tp} = & (1 + x_{\text{im}})(\delta_{p' p'_A}^l + \delta_{p' p'_A}^o \cos 2\psi_A + \delta_{p' p'_A}^u \sin 2\psi_A) \\ & - (1 - x_{\text{im}})(\delta_{p' p'_B}^l - \delta_{p' p'_B}^o \cos 2\psi_B - \delta_{p' p'_B}^u \sin 2\psi_B) \end{aligned} \quad (51)$$

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,

$$\mathbf{P}^T \mathbf{N}^{-1} \mathbf{P} \mathbf{m} = \mathbf{P}^T \mathbf{N}^{-1} \mathbf{d}. \quad (52)$$

For a single-horn experiment, i.e., *Planck* LFI, this reduces to a 3×3 matrix that can be inverted for each pixel independently. For the pointing matrix in Eq. (51), 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

$$\mathbf{P}_{\text{am}}^T \mathbf{N}^{-1} \mathbf{P} \mathbf{m} = \mathbf{P}_{\text{am}}^T \mathbf{N}^{-1} \mathbf{d}. \quad (53)$$

This equation can be solved using the BiCG-STAB algorithm for a non-symmetric matrix \mathbf{A} where $\mathbf{A}\mathbf{x} = \mathbf{b}$. 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 $\mathbf{r} \equiv \mathbf{b} - \mathbf{A}\mathbf{x}$ 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.

2.8.3. Transmission imbalance estimation

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

$$d_{t,j} = g_{t,j}[(1 + x_{\text{im},j})s_{t,j}^{\text{tot,A}} - (1 - x_{\text{im},j})s_{t,j}^{\text{tot,B}}] + n_t. \quad (54)$$

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

$$d_{t,j} = g_{t,j}[s_{t,j}^{\text{d}} + x_{\text{im},j}s_{t,j}^{\text{c}}] + n_t. \quad (55)$$

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. (42) and correlated noise estimation of (44), with $\mathbf{r} = \mathbf{d} - \mathbf{gs}^{\text{d}}$,

$$[(\mathbf{gs}^{\text{c}})^T \mathbf{N}^{-1} \mathbf{gs}^{\text{c}}]x_{\text{im}} = (\mathbf{gs}^{\text{c}})^T \mathbf{N}^{-1} \mathbf{r} + (\mathbf{gs}^{\text{c}})^T \mathbf{N}^{-1/2} \boldsymbol{\eta}. \quad (56)$$

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_{im} .

The WMAP procedure, described in 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.

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

2.8.4. Baseline sampling

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

$$\mathbf{d} = \mathbf{GPBM}\mathbf{a} + \mathbf{n} + \mathbf{b}, \quad (57)$$

where \mathbf{b} is the instrumental baseline and \mathbf{n} is the total instrumental noise. As noted above, Commander3 divides the noise into $\mathbf{n} = \mathbf{n}^{\text{w}} + \mathbf{n}^{\text{corr}}$, 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}^{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 in 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 it to make them compatible with

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.

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	9.3	0.3	0.3	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	17.8
Other initialization														13.4
Total initialization														31.2
<i>Gibbs sampling steps (cost per sample)</i>														
Huffman decompression	1.1	2.1	10.5	0.9	0.8	1.0	1.0	1.3	1.3	1.8	1.8	1.8	1.8	27.2
TOD projection (P operation)	0.4	0.9	4.2	2.6	2.6	3.3	3.4	4.3	4.3	6.4	6.3	6.3	6.4	54.0
Sidelobe evaluation	1.0	2.1	7.6	2.9	2.9	3.5	3.5	4.7	4.8	7.0	6.9	6.9	6.9	60.7
Orbital dipole	0.9	1.9	7.1	1.3	1.3	1.7	1.7	2.2	2.3	3.4	3.3	3.3	3.3	33.7
Gain sampling	0.5	0.8	1.9	0.8	0.8	0.5	0.5	0.9	0.9	0.7	0.7	0.7	0.7	10.4
1 Hz spike sampling	0.3	0.4	1.6											2.4
Correlated noise sampling	2.0	4.0	21.7	2.8	2.9	3.3	3.6	5.1	5.4	8.0	7.7	7.2	8.5	81.3
Correlated noise PSD sampling	4.8	5.9	1.5	0.2	0.2	0.3	0.3	0.5	0.4	0.7	0.6	0.6	0.7	16.7
TOD binning (P' operation)	0.1	0.1	4.0	0.5	0.5	0.7	0.8	0.8	0.8	1.2	1.2	1.2	1.2	13.1
Mapmaking						6.4	7.0	8.9	8.1	11.1	9.5	14.4	14.3	119.5
Sum of other TOD processing	4.4	8.6	44.4	14.7	4.6	5.1	5.0	9.4	7.7	8.1	6.8	8.6	8.7	136.1
TOD processing cost per sample	15.5	26.8	104.5	23.0	24.1	27.6	27.9	40.3	37.4	51.7	50.6	51.9	54.6	535.9
Amplitude sampling														14.0
Spectral index sampling														25.5
Total cost per sample														581.2

Commander3 in Sect. 3.2, then describe the computational requirements in Sect. 3.3.

3.1. Publicly available WMAP products

The full *WMAP* dataset is hosted at the Legacy Archive for Microwave Background Data Analysis (LAMBDA).⁴ 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.⁵ 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 pre-processing 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 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. 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_{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 (as well as the final chunk) 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 parameters. 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 may be more suspect than stable

⁴ https://lambda.gsfc.nasa.gov/product/wmap/dr5/m_products.html

⁵ https://lambda.gsfc.nasa.gov/product/wmap/dr5/tod_info.html

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 analysis plan

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 ~ 20 times. As discussed in Sec. 2.8.2, 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, 250 *V*-band samples, and 125 *W*-band 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

4.1. Trace plots and correlations

To illustrate the dependence of the goodness-of-fit on the noise model, we inspect the 50th TOD segment, corresponding to MJDs 52285.2–52290.6, as a function of Gibbs iteration, in Fig. 4. This is one of the worst-fitting TOD segments of the entire mission, with a reduced relative χ^2 of -7.5 , equivalent to $\chi^2/n = 0.993$. The line plots demonstrate a strong correlation between the noise parameters and the χ^2 , while the gain itself is almost completely uncorrelated with the variations in the χ^2 . As σ_0 is not formally sampled in the Gibbs chain, it is weakly dependent on f_{knee} and α , making it more likely that it is the driver of the correlations in this figure.

4.2. Gain and baselines

To compare the calibrated TODs from *WMAP* versus COSMOGLOBE, it is important to look at the *WMAP* gain model,

$$g = \alpha \frac{\bar{V} - V_0 - \beta(T_{\text{RXB}} - 290 \text{ K})}{T_{\text{FPA}} - T_0} + (m\Delta t + c), \quad (58)$$

where α , V_0 , β , T_0 , m , and c are fit to a constant value across the mission for each radiometer. \bar{V} represents the radio frequency bias powers per detector, and T_{RXB} and T_{FPA} are the receiver box and focal plane assembly temperatures, which are recorded every 23.04 s. Evaluating the model as a function of T_{RXB} and T_{FPA} requires finding the housekeeping data for the thermistor that was physically closest to the relevant radiometer's focal plane on the satellite. As this requires detailed technical information about the specifications of the satellite's schematics layout that can easily be misunderstood, we do not attempt to reproduce the gain model given in Eq. (58) in this work. Although we are unable to reproduce the exact gain model parametrized in Greason et al. (2012), the 23.04 s time dependence of the gain model on housekeeping data is a plausible explanation for the time-dependent noise variation in the different calibrated data solutions.

As reported in Hinshaw et al. (2007), the calibrated data archive has been calibrated using the procedure listed above, with a baseline subtracted each hour and the sidelobe subtracted. Figure 6 shows the COSMOGLOBE timestream $\mathbf{d}/g - \mathbf{s}_{\text{sl}} - \mathbf{b}$ with the *WMAP* delivered calibrated signal subtracted. 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. The differences of order $50 \mu\text{K}$ correspond to sky brightness of order 10 mK , 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.

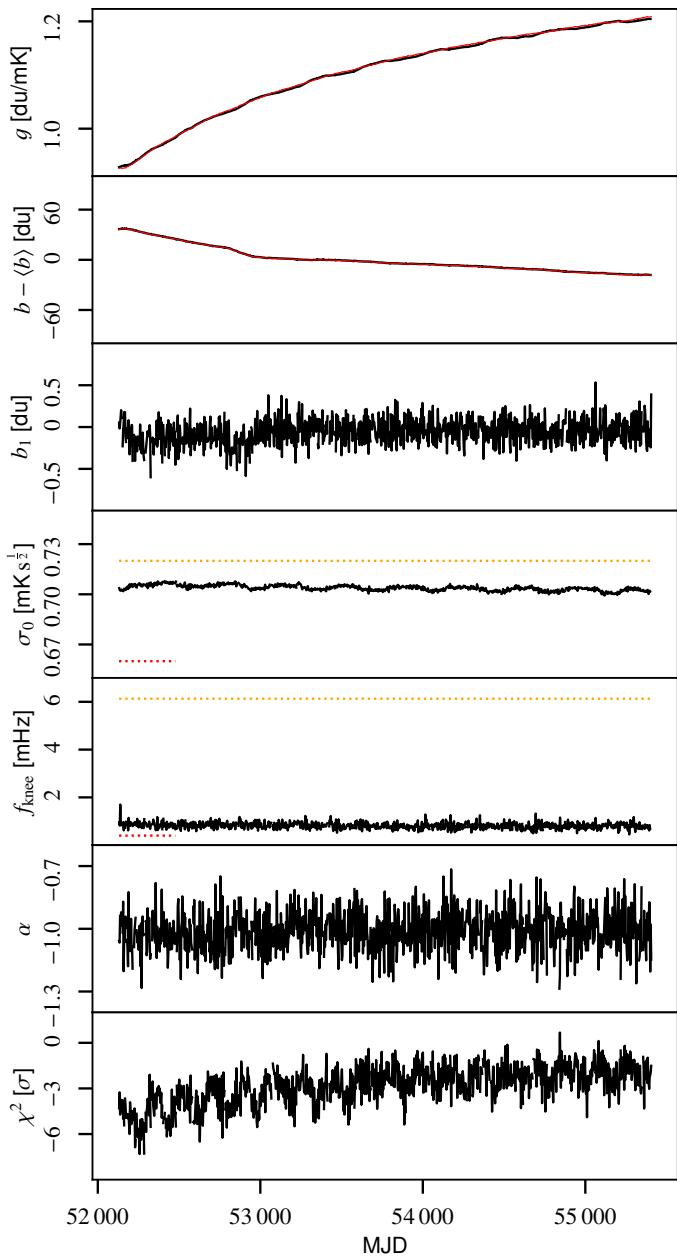


Fig. 3. Overview of K113. The red solid lines in first and second panel are the regressed gain and baseline from *WMAP9*, while the black lines in all panels are samples from the COSMOGLOBE Gibbs chain. The red dashed and yellow dashed lines are reported σ_0 and f_{knee} values from the first-year *WMAP* data analysis and GSFC measurements, respectively.

On longer timescales, as displayed in Figure 7, the most prominent feature is a varying signal of amplitude 0.2 mK. This likely due to the hourly baseline subtraction mentioned above, which contrasts with the COSMOGLOBE approach of assigning a linear baseline solution for the entire scan. The variations are commensurate with correlated noise, which for K113 has $f_{\text{knee}} \sim 0.5$ mHz, corresponding to a little over half an hour. Therefore, the hourlong baseline subtraction essentially acts as a destriper, removing an estimate of the correlated noise. To test this hypothesis, we plot a realization of correlated noise generated by Commander, and find that the signals are very similar, both in amplitude and morphology.

We also compare the gain and baseline solutions throughout the course of the mission in Fig. 3. To recover the *WMAP9* gain

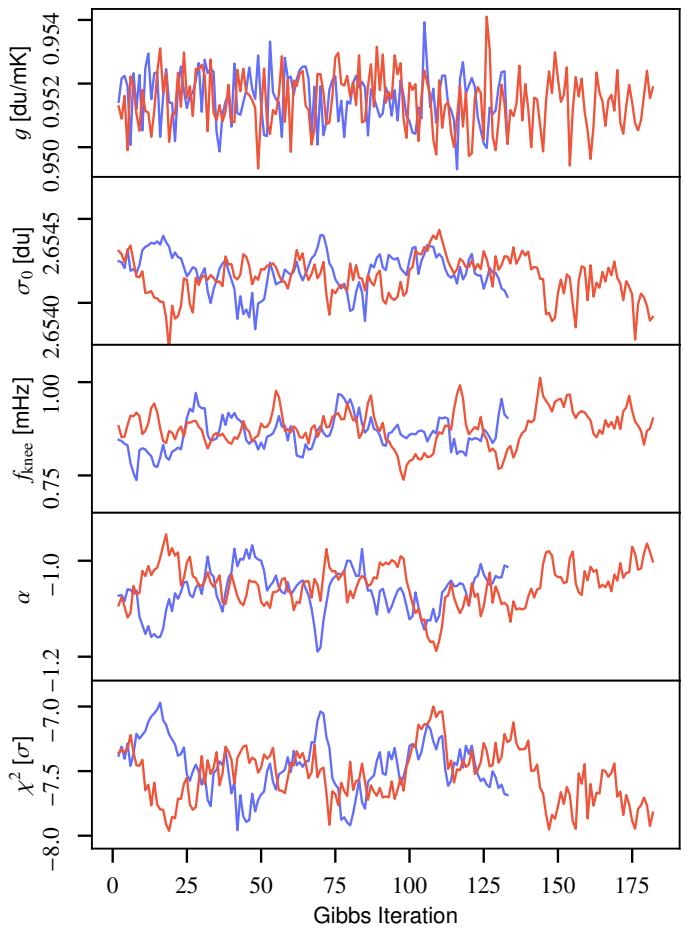


Fig. 4. Subset of K113 Gibbs samples for both chains, arbitrarily colored red and blue. The parameters correspond to MJDs 52285.2–52290.6.

solution, we directly compare the uncalibrated *WMAP* data with the calibrated *WMAP* data with a far sidelobe contribution convolved with the delivered *WMAP9* DA maps. We find that the calibrated and uncalibrated data can be related by

$$d_t^{\text{raw}} = g(d_t^{\text{cal}} + s_t^{\text{sl}}) + \sum_{i=0}^3 c_i(t - t_0)^i, \quad (59)$$

where the second term is a cubic polynomial with coefficients c_i referenced to the time at the beginning of the scan t_0 . To calculate s_t^{sl} , we convolve the *WMAP* far sidelobes with *WMAP9* frequency maps with the Solar dipole from Hinshaw et al. (2009) added back in. We find that d^{raw} is consistent with the expression on the right at the level of < 0.1 du for all radiometers, suggesting that this estimate of g and the baseline c_0 is a good approximation of the *WMAP9* calibration solution. An initial estimate using a linear baseline gave an unacceptably poor fit. Given that Eq. (2) of Jarosik et al. (2003a) employed a cubic baseline fit while fitting for transmission imbalance parameters, it is reasonable to assume that the official calibrated archive was created using a similar procedure.

The morphological characteristics of the *WMAP9* and COSMOGLOBE gain solutions are similar, with a general trend to increase with time. Both solutions also follow a sinusoidal pattern, corresponding to temperature change due to L2's motion around the Sun (Greason et al. 2012). However, we do find the COSMOGLOBE K-band gain has slightly more oscillatory features



Fig. 5. 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, σ_0 [mK]. The values for each detector are ordered 13, 14, 23, and 24.

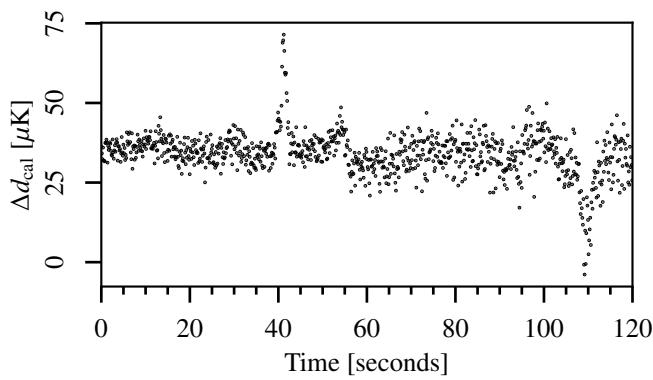


Fig. 6. Difference between the CosmoGlobe $d_{\text{cal}} = d/g - b - s_{\text{sl}}$ and the delivered calibrated TOD from WMAP.

than the WMAP9 solution. In general, the gains are consistent between CosmoGlobe and WMAP9 within 1 %. For completeness, the full gain comparisons can be found in Fig. A.3.

4.3. Transmission imbalance

The transmission imbalance parameters x_{im} are crucial to measure correctly because their misestimation can induce a large polarized signal that is coupled to the Solar dipole (Jarosik et al. 2007; Watts et al. 2022). The uncertainty in x_{im} was quoted as the source of large-scale polarized features in the WMAP9 maps, and a template of this effect was explicitly projected out in the pixel-space polarized covariance matrix.

We find x_{im} values that are largely consistent with the values reported in Bennett et al. (2013), albeit with some outliers. We find in general that the 68 % confidence intervals from CosmoGlobe are smaller than the fiducial values, although we caution

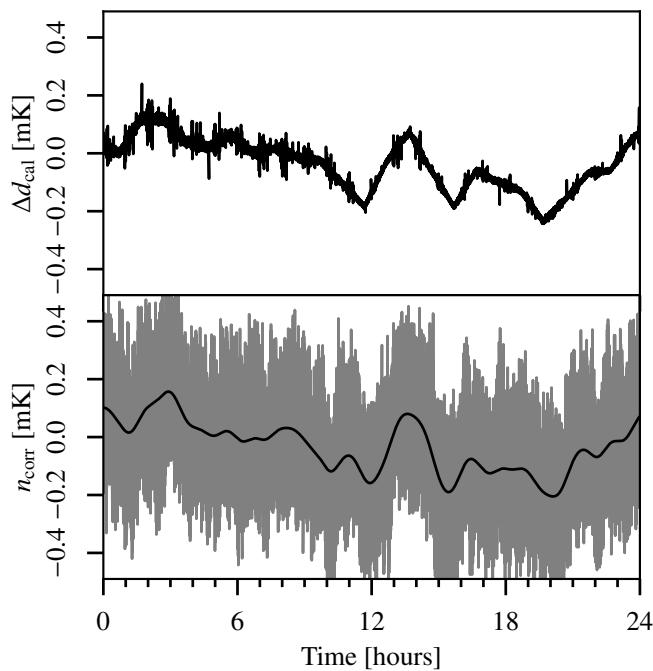


Fig. 7. (top): Difference between the COSMOGLobe $d_{\text{cal}} = d/g - b - s_{\text{sl}}$ and the delivered calibrated TOD from *WMAP*. (Bottom): Raw correlated noise (gray) and smoothed data with Gaussian kernel (black). This shows more clearly the hourly baseline subtraction from the *WMAP* treatment.

against a direct comparison of these values since such different procedures were used for estimating the uncertainties.

4.4. Instrumental noise and goodness-of-fit

The noise fitting, as outlined in Sect. 2.8, inherently depends on the data being fit well by both the sky model and the instrument model. In practice, correlated noise fitting can model any unmodeled signals, so the power spectrum and TODs must be carefully scrutinized before any conclusions can be made about the corresponding maps.

The white noise level in raw du is not strictly sampled, but is estimated conditioned on the instrumental parameters and the sky parameters. However, the calibrated white noise level $\sigma_0[\text{K}] = \sigma_0[du]/g$ does depend on the gain quite directly, which allows us to test the effects of the calibration on the instrument sensitivity itself. The calibrated white noise level follows a biaurnal trend indicative of a system temperature variation, which is to be expected given the radiometer equation

$$\sigma_0[\text{V}] \propto gT_{\text{sys}}. \quad (60)$$

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.

The knee frequencies for each channel lie between the reported values in Jarosik et al. (2003a) for both the Goddard Space Flight Center (GSFC) laboratory measurements and those from the first year of data collection. Nearly all radiometers have constant f_{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 α is around -1 for each radiometer, albeit with high scatter for the lower frequencies. As expected, the uncertainty in α decreases as f_{knee} increases, since there are more datapoints to fit below f_{knee} where the constraining power on α is the strongest.

The most striking feature of the reduced normalized χ^2 is its amplitude and its semiannual periodicity. Given the noise model and data residual, we can evaluate the goodness-of-fit in the form of the relative χ^2 . Here, we find that approximately half of the radiometers have a χ^2 value at least 6σ above or below the expected value. Given perfect Gaussian residuals, we would expect the reduced sum of squares to be $n_{\text{TOD}} = 2^N$ and be within $\sqrt{2n_{\text{TOD}}} = 2^{(N+1)/2}$ 68 % of the time. For a typical W-band scan of length $n_{\text{TOD}} = 2^{22}$, a 10σ model failure corresponds to $\chi^2/n_{\text{TOD}} = 1.003$. Therefore, it is exceedingly difficult to look at any given *WMAP* scan in the time domain and identify a model failure. In power spectrum space, i.e., in Fig. 8, the data are still characterized well at all scales, despite this scan having a χ^2 that is 7σ above the expectation value.

Only with aggressive smoothing, as in Fig. 9, does the model failure become apparent at frequencies 1–10 Hz. 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 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^α . 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_{TOD} giving a higher uncertainty on the relative χ^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.

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-

Table 3. Summary of noise properties.

Radiometer ..	Diode	Sensitivity, σ_0 (mK \sqrt{s})			Knee frequency, f_{knee} (mHz)			Slope, α
		GSFC	WMAP	CG/ $\sqrt{2}$	GSFC	WMAP	CG/ $\sqrt{2}$	
K11	1	0.72	0.66	0.704 ± 0.002	6.13	0.4	0.82 ± 0.20	-1.01 ± 0.10
	2			0.708 ± 0.003			0.63 ± 0.14	-0.95 ± 0.10
K12	1	0.87	0.75	0.796 ± 0.004	5.37	0.51	0.42 ± 0.19	-0.93 ± 0.12
	2			0.780 ± 0.005			0.71 ± 0.15	-1.02 ± 0.10
Ka11	1	0.75	0.71	0.788 ± 0.001	1.66	0.71	1.20 ± 0.22	-1.02 ± 0.09
	2			0.777 ± 0.001			1.19 ± 0.22	-1.02 ± 0.09
Ka12	1	0.77	0.72	0.788 ± 0.003	1.29	0.32	0.62 ± 0.16	-0.99 ± 0.11
	2			0.784 ± 0.001			0.63 ± 0.13	-1.01 ± 0.11
Q11	1	0.99	0.92	0.998 ± 0.002	3.21	1.09	1.06 ± 0.16	-1.09 ± 0.09
	2			0.992 ± 0.002			1.06 ± 0.16	-1.10 ± 0.09
Q12	1	0.95	1.02	1.159 ± 0.007	3.13	0.35	0.45 ± 0.47	-0.98 ± 0.11
	2			1.146 ± 0.007			0.83 ± 0.14	-1.00 ± 0.09
Q21	1	0.89	0.85	0.908 ± 0.002	1.92	5.76	2.88 ± 0.37	-1.10 ± 0.07
	2			0.906 ± 0.002			3.22 ± 0.56	-1.10 ± 0.06
Q22	1	1.04	0.99	1.074 ± 0.004	4.61	8.62	3.95 ± 0.54	-1.11 ± 0.06
	2			1.064 ± 0.003			4.05 ± 0.64	-1.11 ± 0.06
V11	1	1.25	1.22	1.551 ± 0.003	2.56	0.09	1.27 ± 0.15	-0.90 ± 0.06
	2			1.539 ± 0.003			1.19 ± 0.14	-0.89 ± 0.06
V12	1	1.07	1.11	1.398 ± 0.002	4.49	1.41	2.11 ± 0.20	-0.97 ± 0.05
	2			1.432 ± 0.002			1.88 ± 0.17	-0.96 ± 0.05
V21	1	1.01	0.97	1.241 ± 0.298	2.43	0.88	1.50 ± 0.24	-0.95 ± 0.07
	2			1.217 ± 0.294			1.60 ± 0.26	-0.97 ± 0.06
V22	1	1.13	1.1	1.443 ± 0.300	3.06	8.35	4.01 ± 0.85	-1.00 ± 0.08
	2			1.415 ± 0.316			3.08 ± 0.65	-1.01 ± 0.08
W11	1	1.18	1.35	1.938 ± 0.005	16.2	7.88	5.59 ± 0.53	-0.94 ± 0.05
	2			1.895 ± 0.005			8.99 ± 0.85	-0.95 ± 0.04
W12	1	1.41	1.61	2.301 ± 0.005	15.1	0.66	3.91 ± 0.42	-0.89 ± 0.05
	2			2.345 ± 0.006			4.81 ± 0.53	-0.89 ± 0.05
W21	1	1.38	1.61	2.225 ± 0.007	1.76	9.02	13.57 ± 1.47	-0.89 ± 0.03
	2			2.292 ± 0.006			5.06 ± 0.95	-0.93 ± 0.05
W22	1	1.44	1.72	2.291 ± 0.006	0.77	7.47	3.02 ± 0.53	-0.98 ± 0.05
	2			2.232 ± 0.007			7.26 ± 1.05	-0.95 ± 0.04
W31	1	1.47	1.65	2.328 ± 0.005	1.84	0.93	1.30 ± 0.46	-0.99 ± 0.07
	2			2.322 ± 0.006			1.97 ± 0.28	-0.98 ± 0.06
W32	1	1.69	1.86	2.707 ± 0.015	2.39	0.28	1.59 ± 0.29	-0.98 ± 0.07
	2			2.579 ± 0.015			1.40 ± 0.39	-1.00 ± 0.07
W41	1	1.6	1.71	2.519 ± 0.010	8.46	46.5	26.81 ± 1.83	-0.92 ± 0.04
	2			2.479 ± 0.009			24.75 ± 1.63	-0.92 ± 0.04
W42	1	1.43	1.65	2.221 ± 0.017	5.31	26.0	16.10 ± 1.09	-0.94 ± 0.04
	2			2.202 ± 0.015			17.11 ± 1.19	-0.94 ± 0.04

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. 10, 11, 12, 13, and 14, respectively, given in μK_{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° 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 param-

eters. The top row of Fig. 15 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° 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.

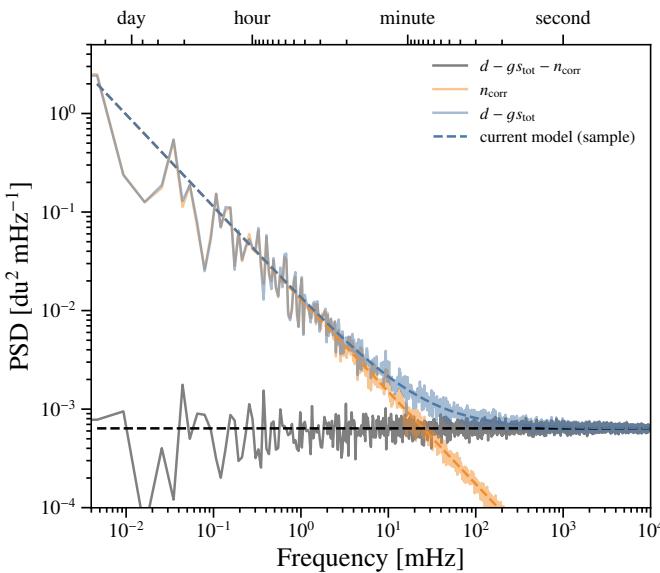


Fig. 8. PSD for radiometer W413 that spans MJDs 52252.3–52254.8. The power spectrum of the blue line corresponds to the residual, while the gray line is the residual with a correlated noise realization removed.

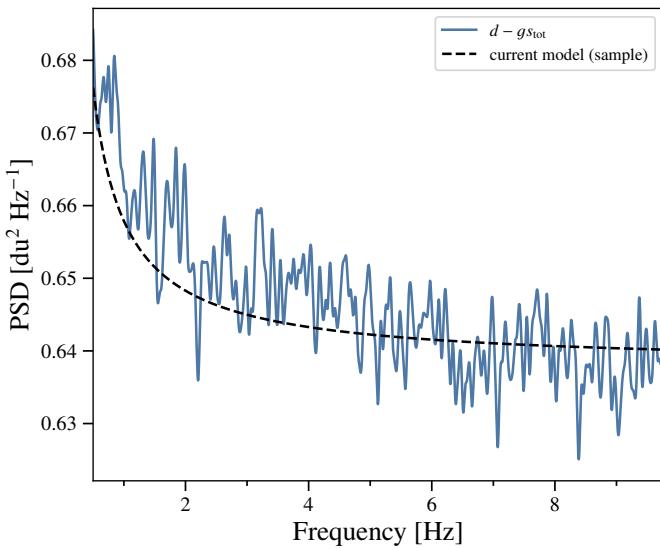


Fig. 9. PSD for radiometer W413 that spans MJDs 52252.3–52254.8. The black dashed line is a sample of the theoretical PSD, while the blue line is the smoothed residual power spectrum.

While the maps in the top row of Fig. 15 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. 16. 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.

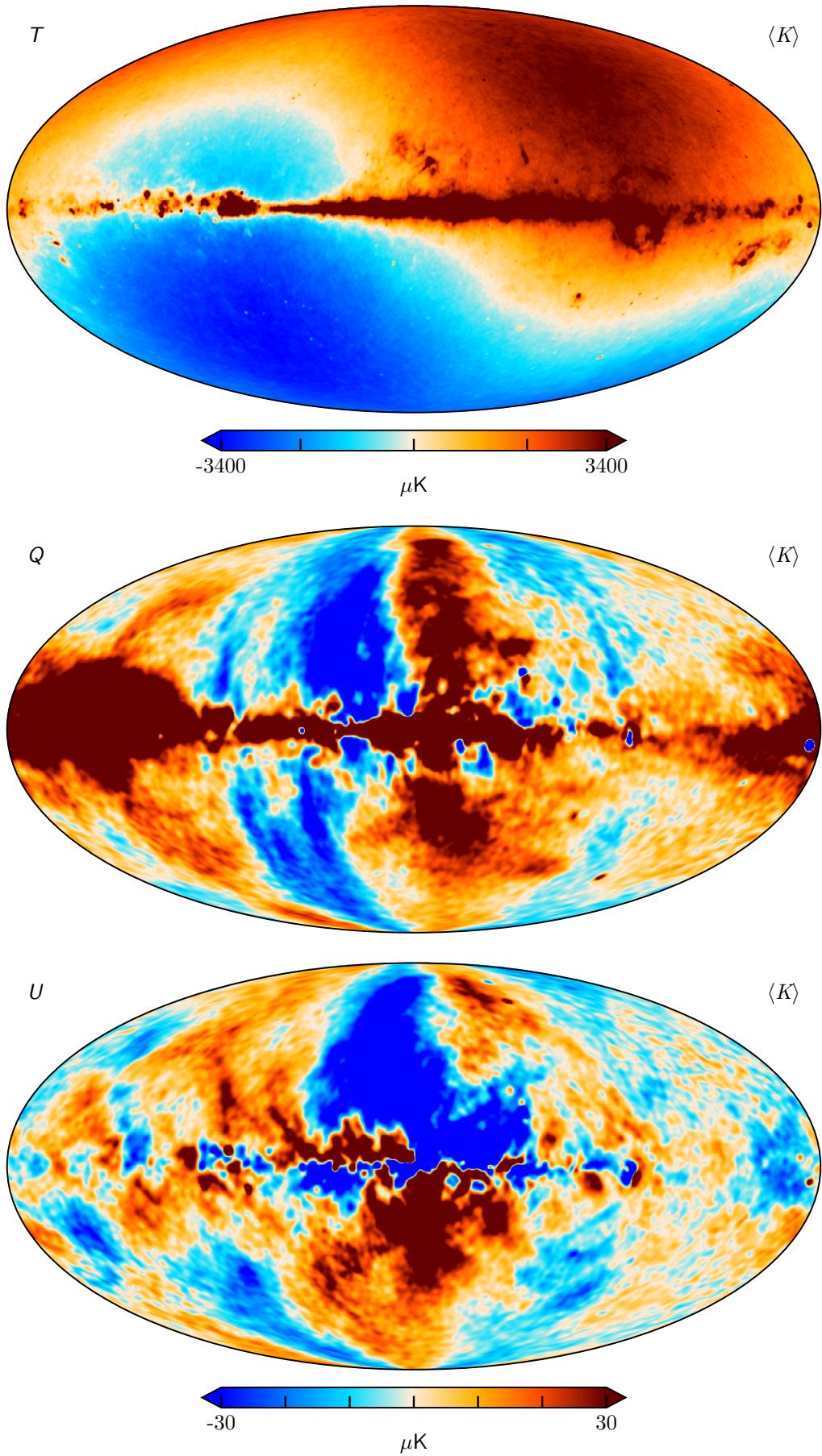
We also consider the spatial structure of the TOD corrections in pixel space, as shown in Fig. 17. This corresponds to the TOD objects in Fig. 1 binned into Stokes *I*, *Q*, and *U* maps. Note that the dynamic range of each of these figures varies by two orders of magnitude, indicating that some components 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. 18, 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_ℓ^{EE} and C_ℓ^{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. 19. 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_ℓ power spectra using the NaMaster (Alonso et al. 2019)⁶ `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. 19. 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.

⁶ <https://github.com/LSSTDESC/NaMaster>

**Fig. 10.** K-band

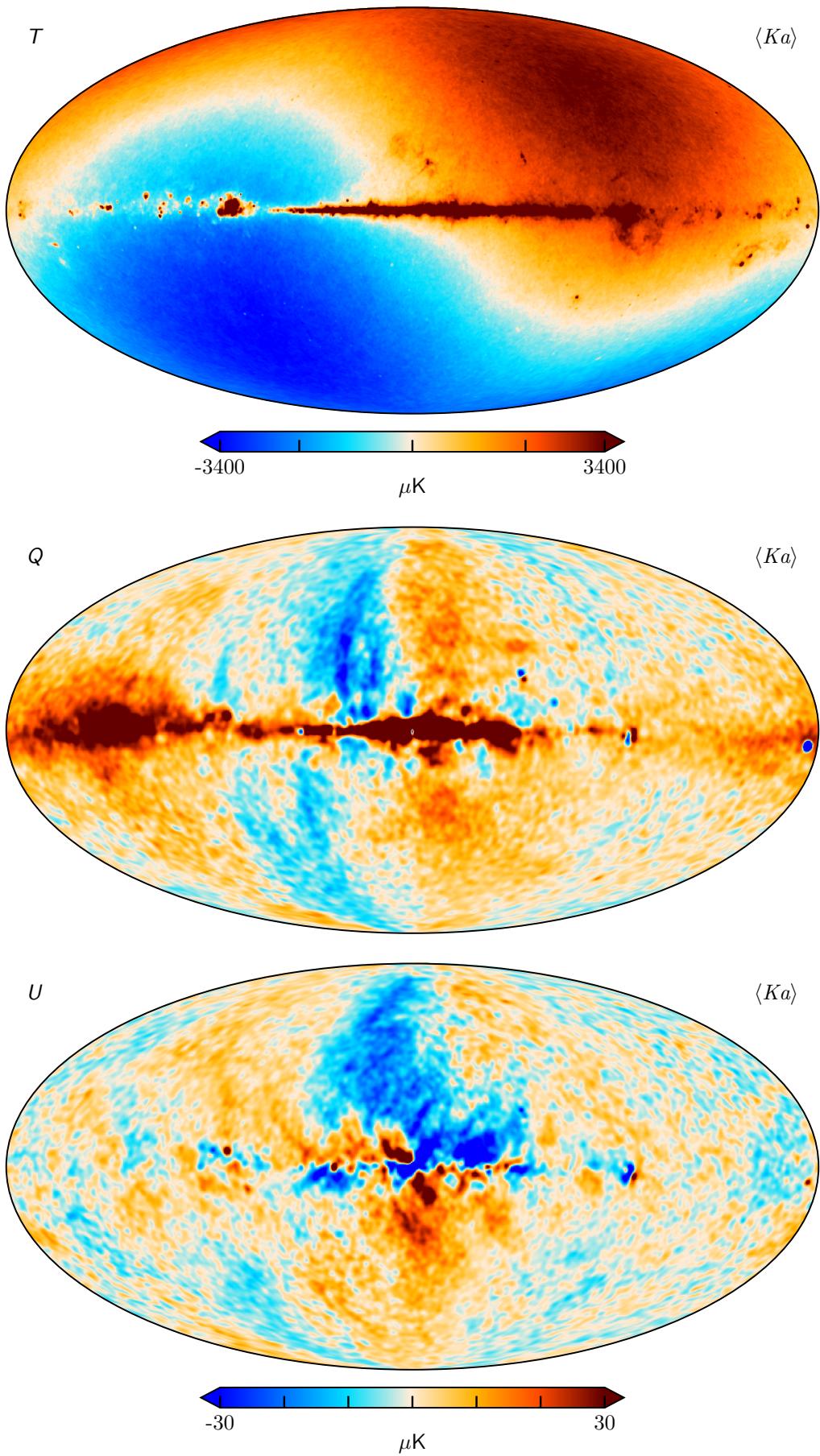


Fig. 11. Ka-band

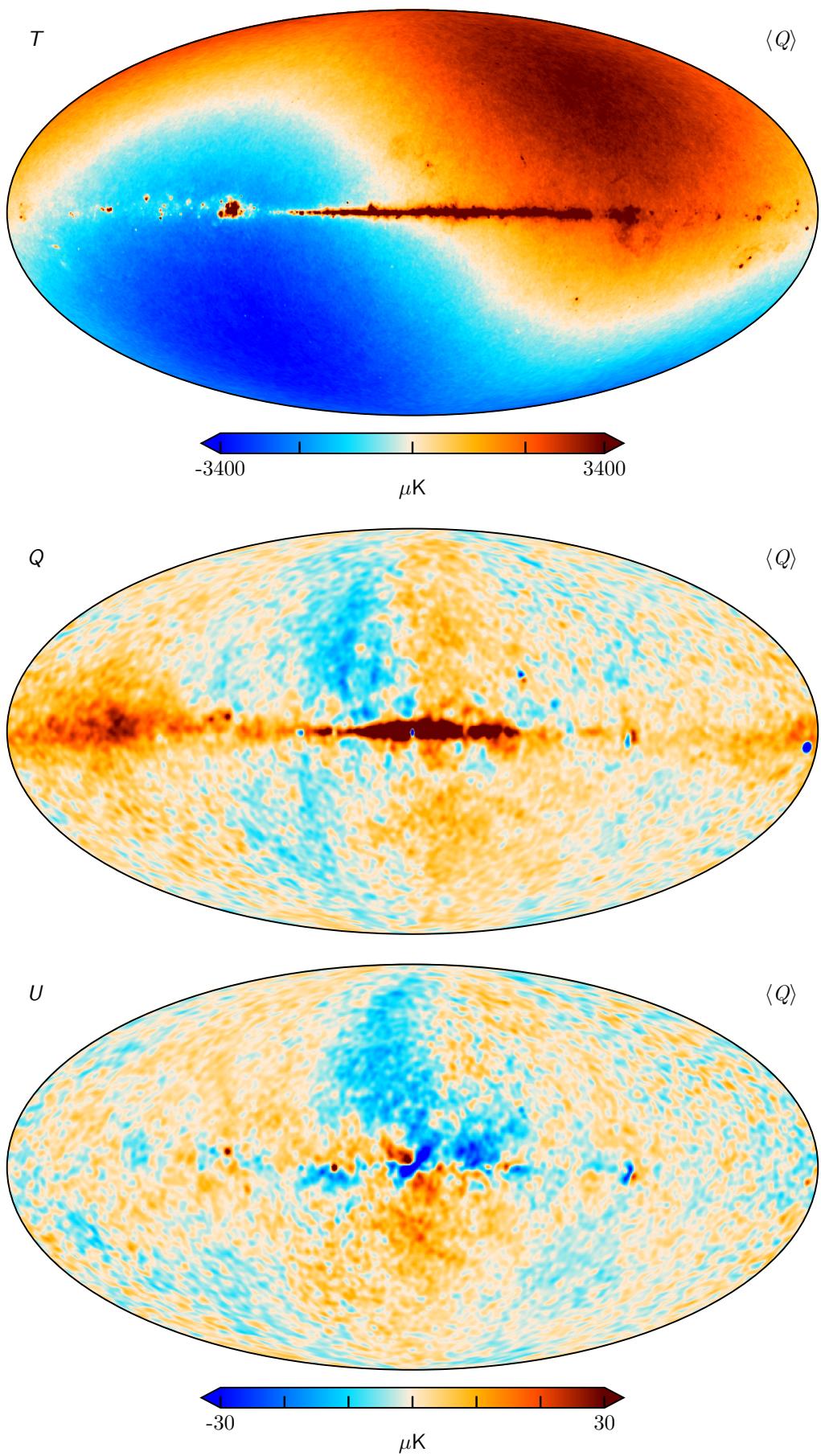


Fig. 12. Q-band

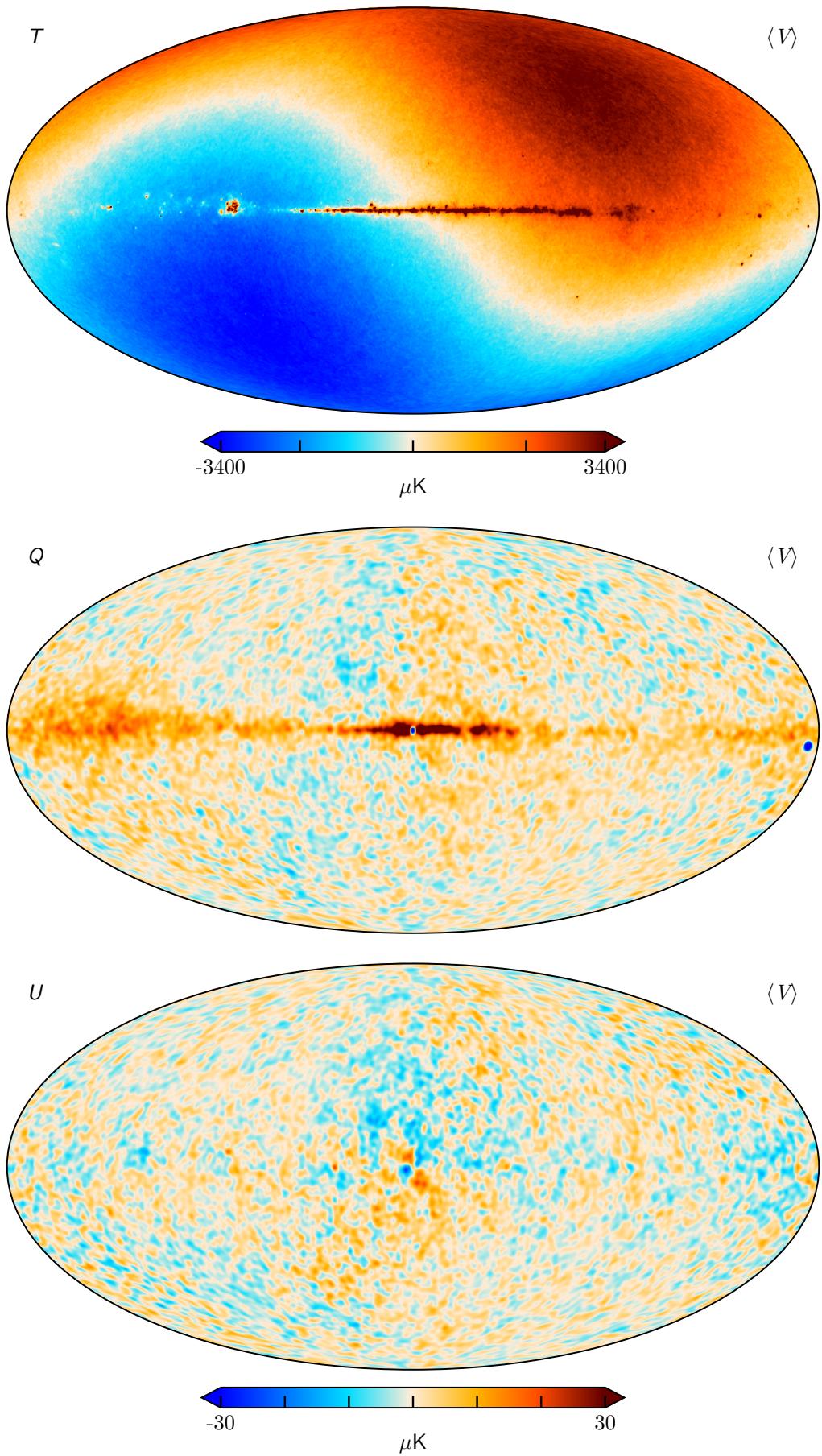


Fig. 13. V-band

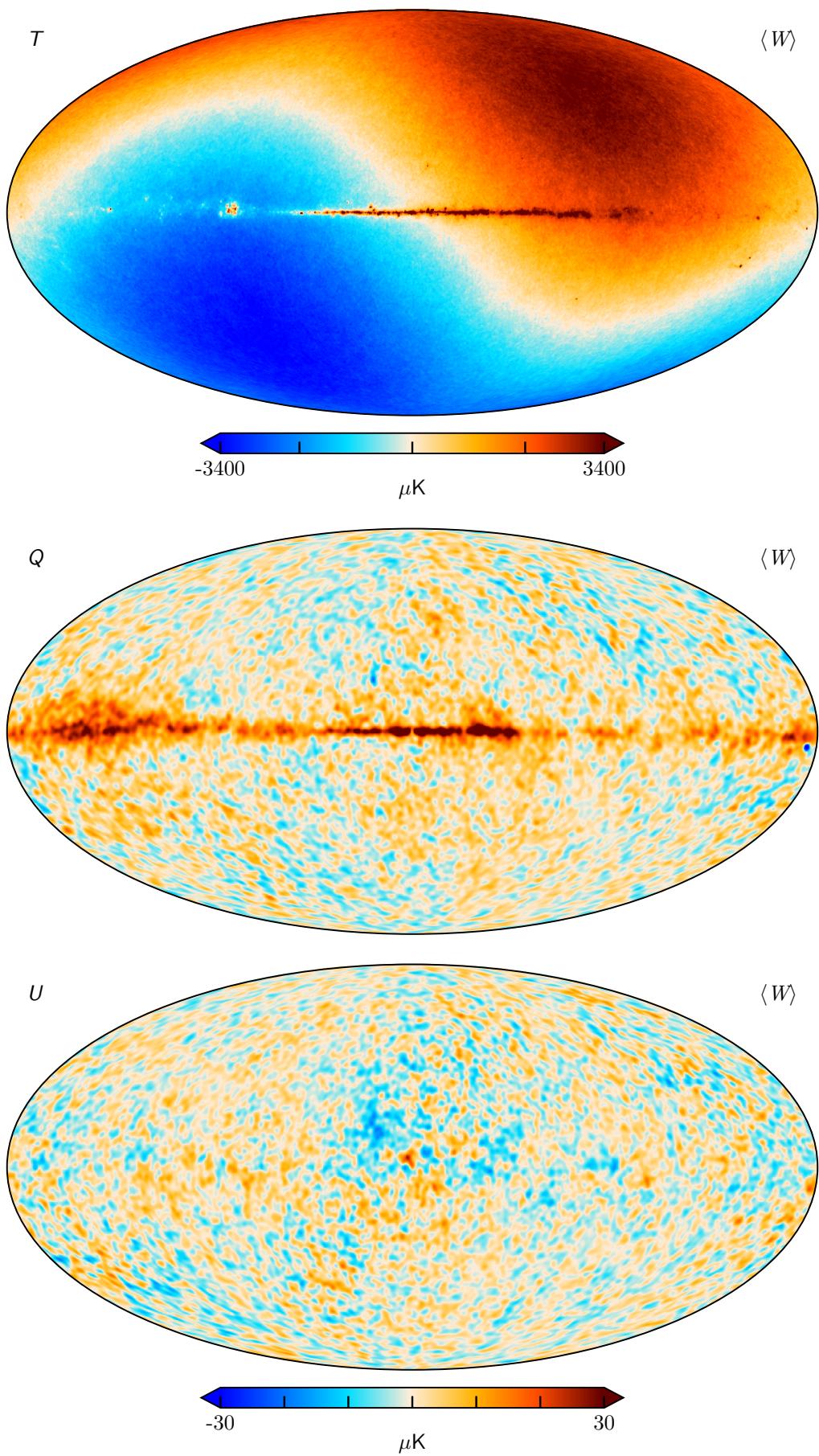


Fig. 14. W-band

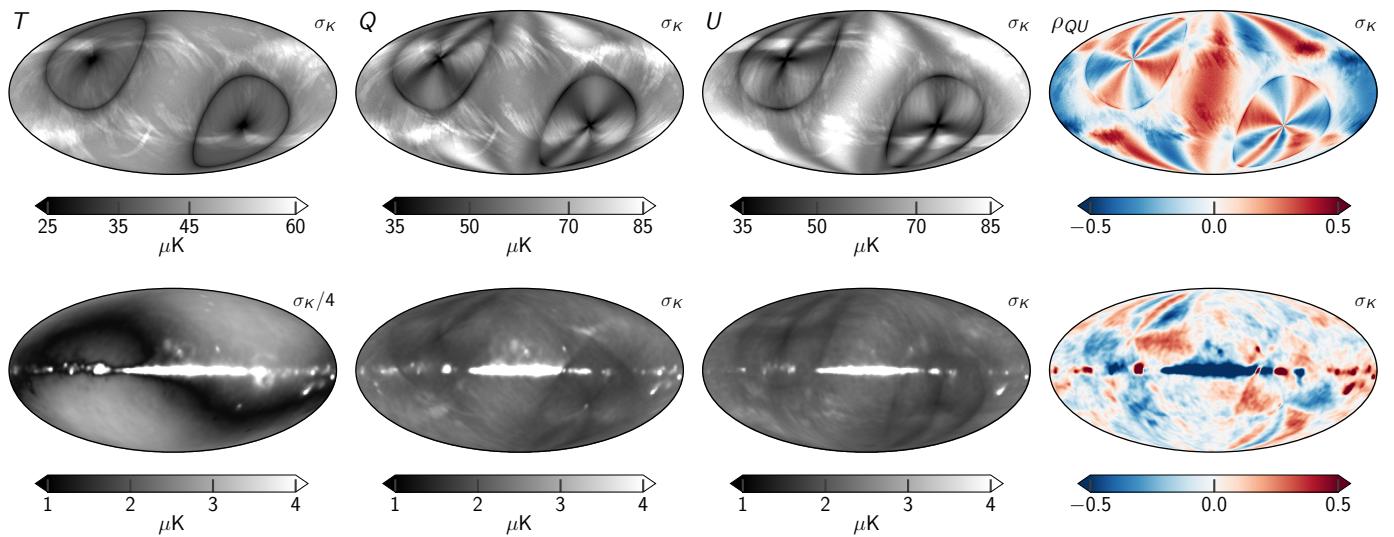


Fig. 15. Posterior variation maps for K-band. Columns show the Stokes parameters and the correlation coefficient between Q and U , while the rows show (top) the white noise rms per pixel and (bottom) the posterior standard deviation. The rms maps are unsmoothed, while the standard deviations have been smoothed to 7° .

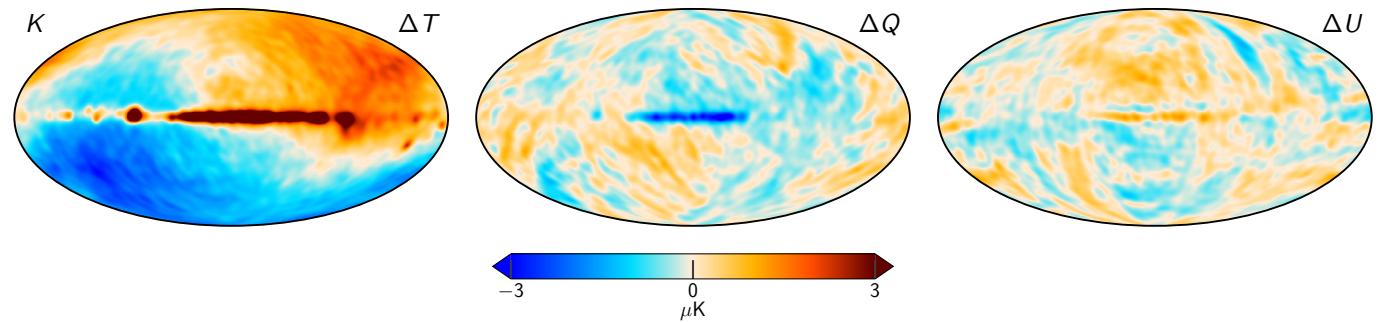


Fig. 16. Difference between two K-band Gibbs samples, smoothed to 7° .

The E -mode power spectra, displayed in the second column of Fig. 19, 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. 19, should follow the same pattern as in the E -modes, but with foregrounds reduced by a factor of $\simeq 2$ (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 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. 20. 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 μK 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 8.3. This can be attributed to the absolute calibration differences between the two pipelines. In particular, Bennett et al. (2013) estimates 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. 2.8.1, 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

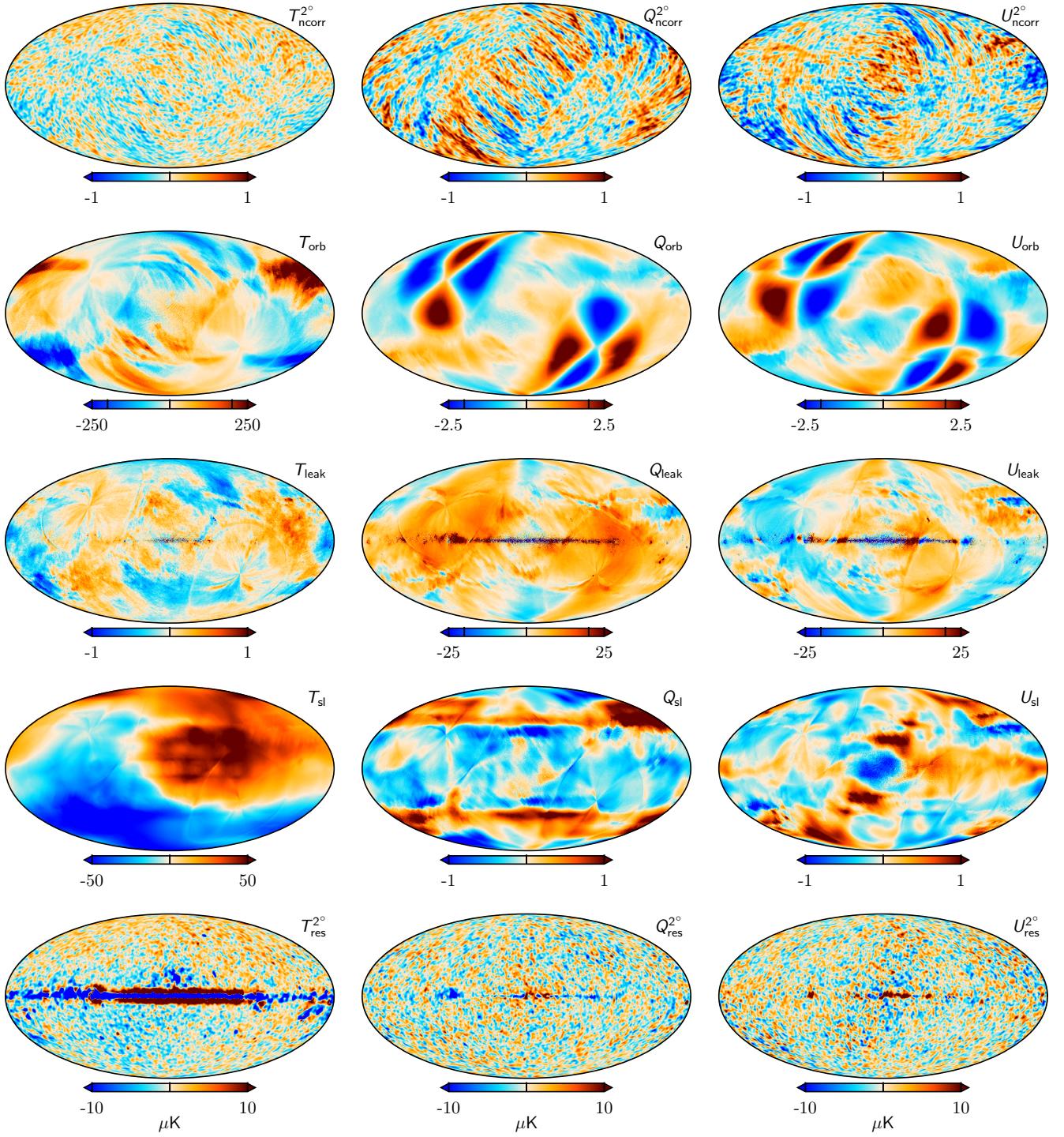


Fig. 17. TOD corrections for K -band for a single Gibbs sample, projected into maps. Columns show Stokes T , Q , and U parameters. Rows show, from top to bottom, 1) correlated noise; 2) the orbital dipole; 3) bandpass mismatch leakage; and 4) sidelobe corrections. The bottom row shows the residual obtained 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° Gaussian beam.

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 in 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 func-

tional 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

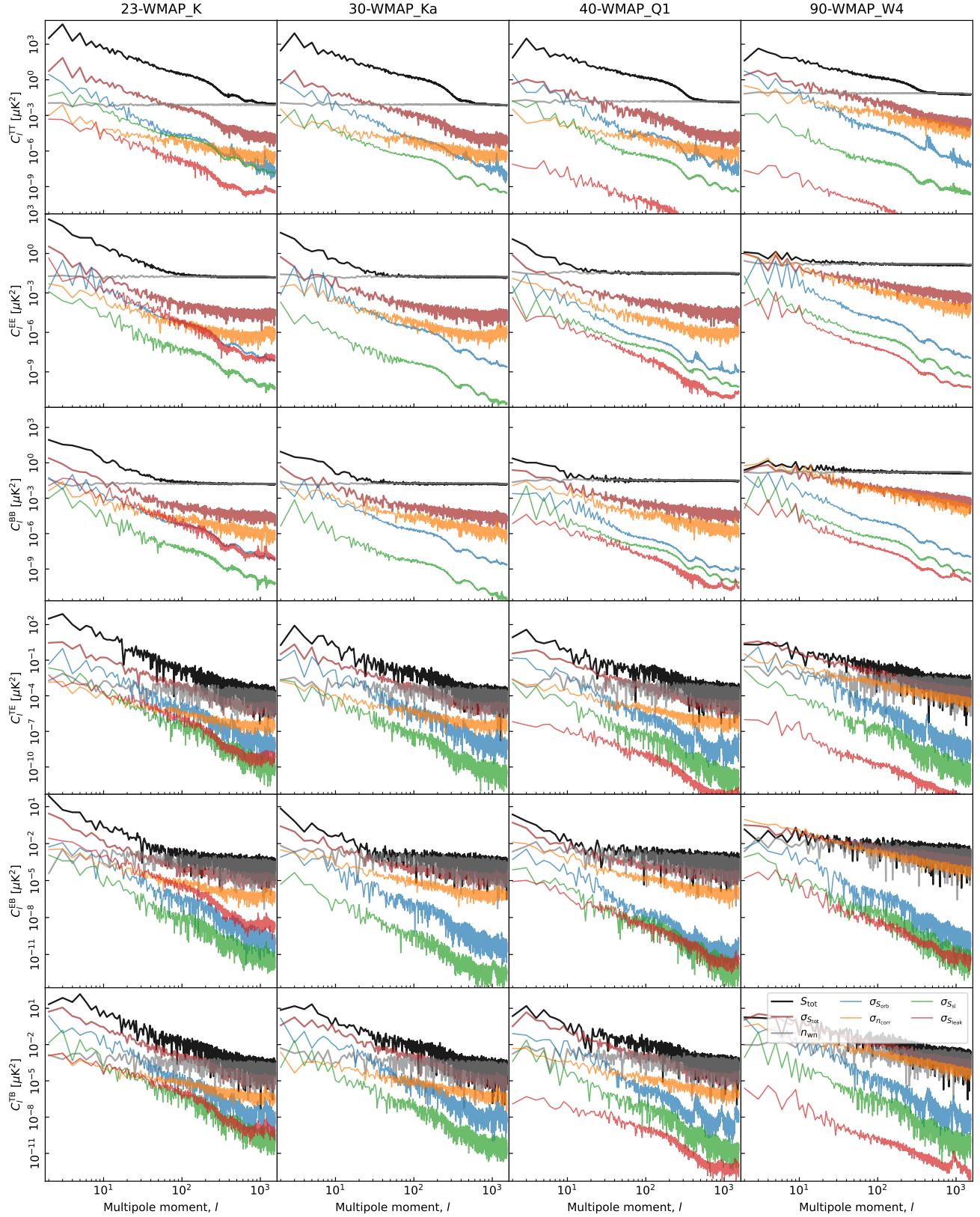


Fig. 18. Pseudo-spectrum standard deviation for each instrumental systematic correction shown in Fig. 17 (thin colored lines). For comparison, thick black lines show spectra for the full coadded frequency map; thick red lines show the standard deviation of the same (i.e., the full systematic uncertainty); 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.

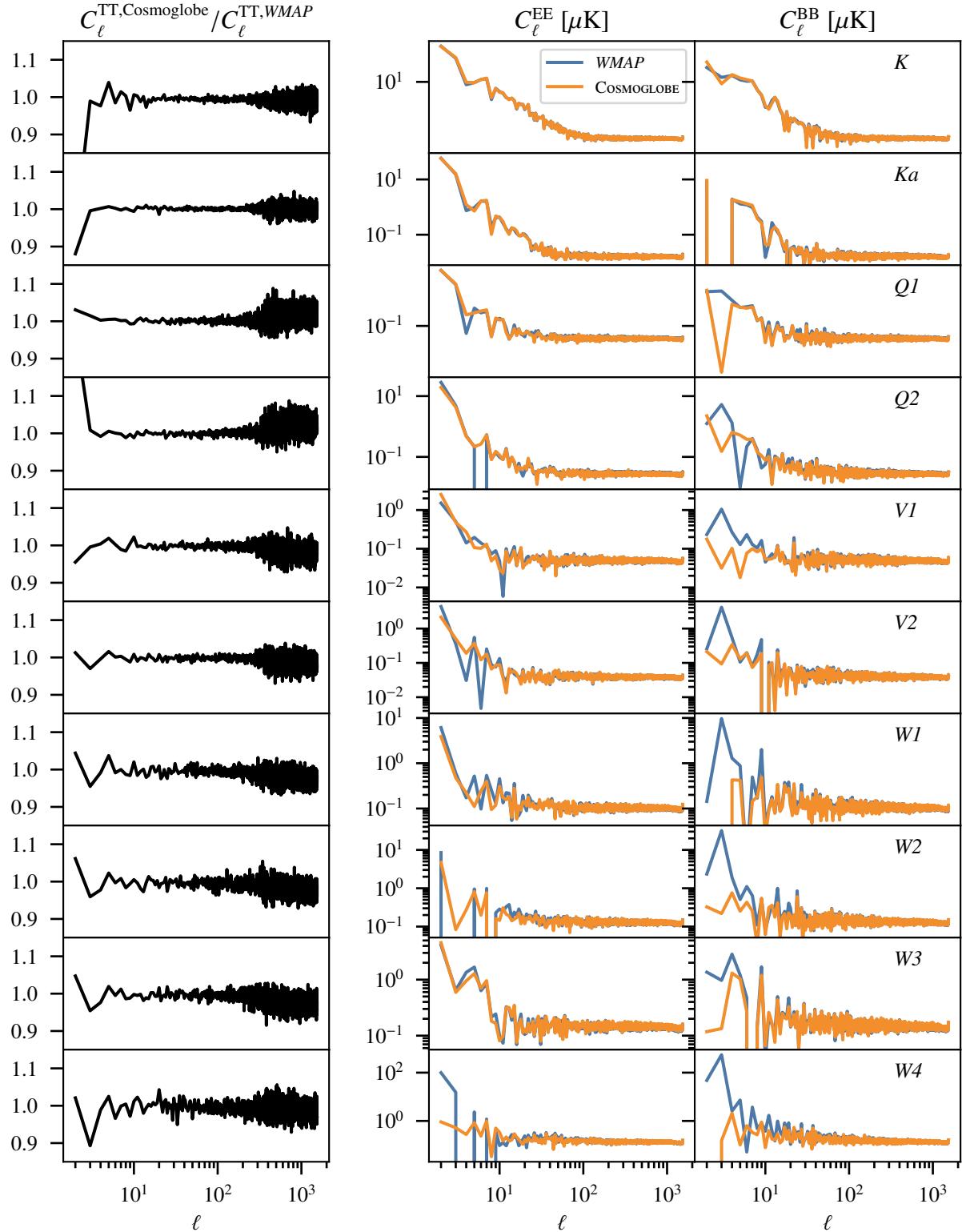


Fig. 19. Comparison of the C_{ℓ}^{TT} , C_{ℓ}^{EE} , and C_{ℓ}^{BB} from WMAP9 and COSMOGLOBE. Each row corresponds to a different DA, with frequency increasing from top to bottom. (left): ratio of C_{ℓ}^{TT} from COSMOGLOBE compared to WMAP9. (middle/right): $C_{\ell}^{EE/BB}$ power spectra with WMAP9 in blue and COSMOGLOBE in orange.

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 sidelobe, and the Solar dipole, as briefly described in Sect. 2.2. The scale of these effect is most pronounced in the W-band polarization results, where we see the largest differences between the two processing pipelines.

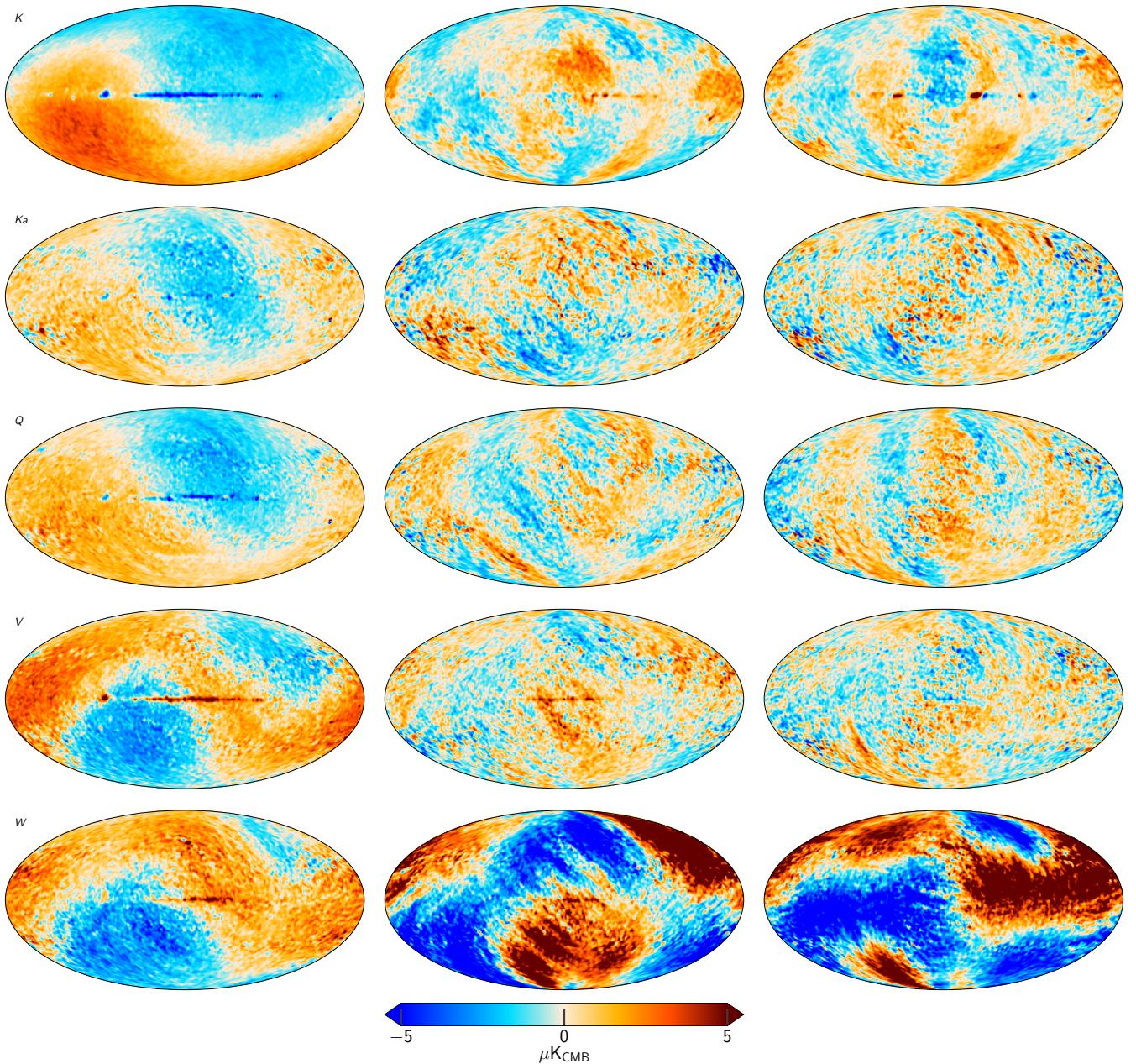


Fig. 20. Difference maps between the COSMOGLobe and 9-year *WMAP* frequency maps. Columns show Stokes T , Q , and U parameter maps, while rows show K -, Ka -, Q -, V -, and W -band maps. The maps are all smoothed with a 2° FWHM Gaussian beam.

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. 20, particularly in polarization, are due to effects in the *WMAP* maps that cannot be constrained by the *WMAP* observation strategy. 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.

Table 4. Difference map χ^2 statistics.

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

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. 2.2, 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

Table 5. 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

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. 21. 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 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- ℓ *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. 22. 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 23 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 BEYONDPLANCK 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. 23 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. 21. 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. 23. 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. 23 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. 2022b).

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.

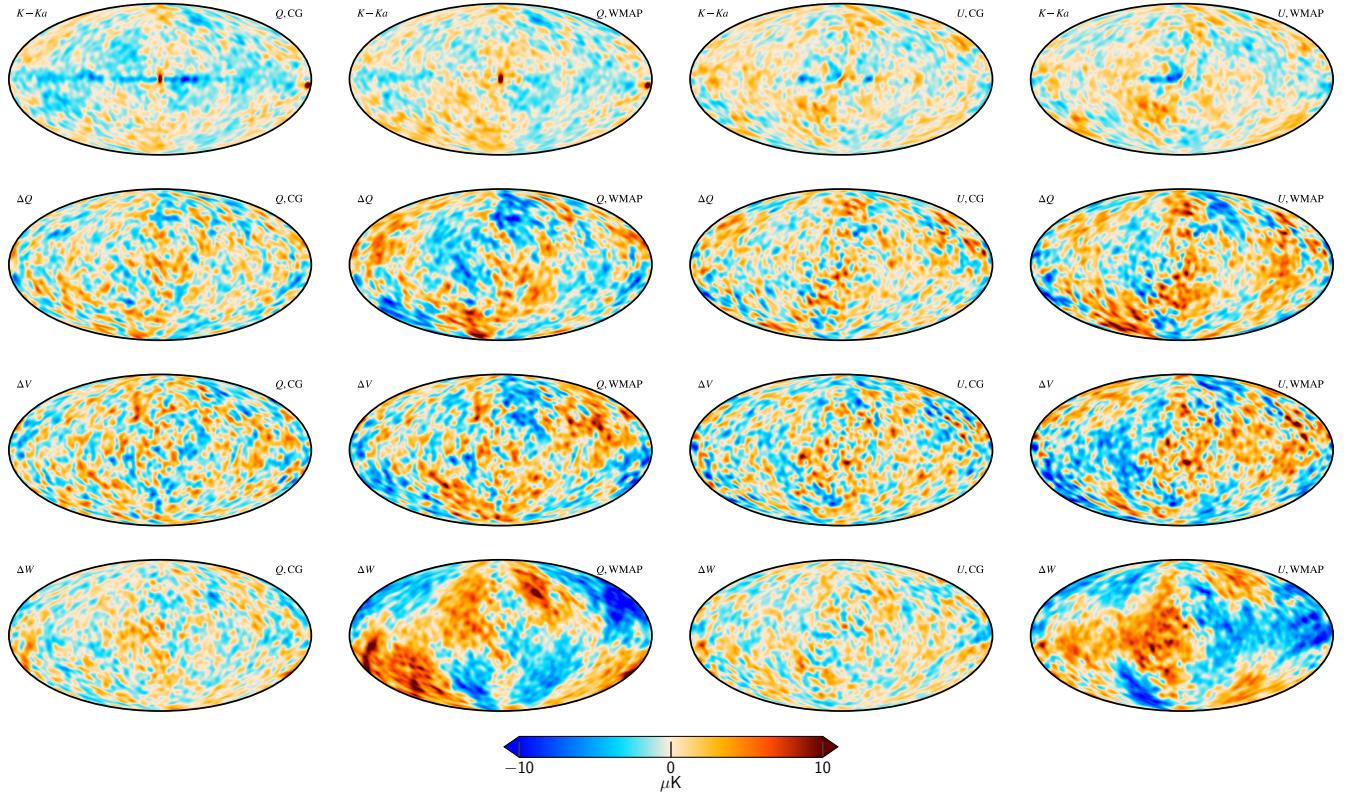


Fig. 21. Internal WMAP difference maps, smoothed by 10° . The two left columns are Stokes Q , and the two right columns are Stokes U , with the COSMOGLOBE and WMAP9 maps alternating between columns. The top to bottom rows are difference maps in increasing frequency.

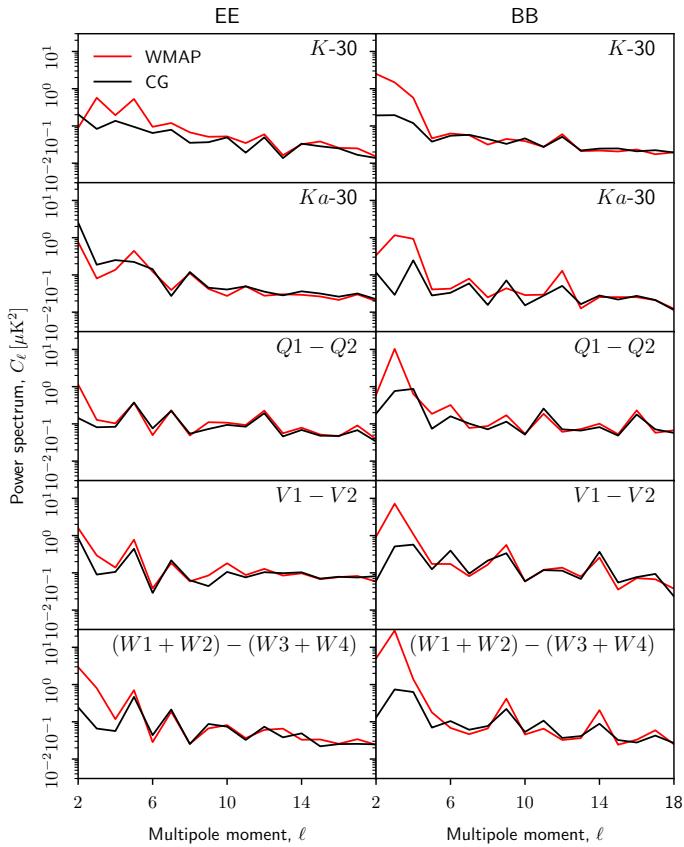


Fig. 22. Full sky half-difference spectra. The red lines are the power spectra of the WMAP9 difference maps, while the black lines are the same for the reprocessed COSMOGLOBE maps.

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 here are similar in quality to the results presented in [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 τ 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 cou-

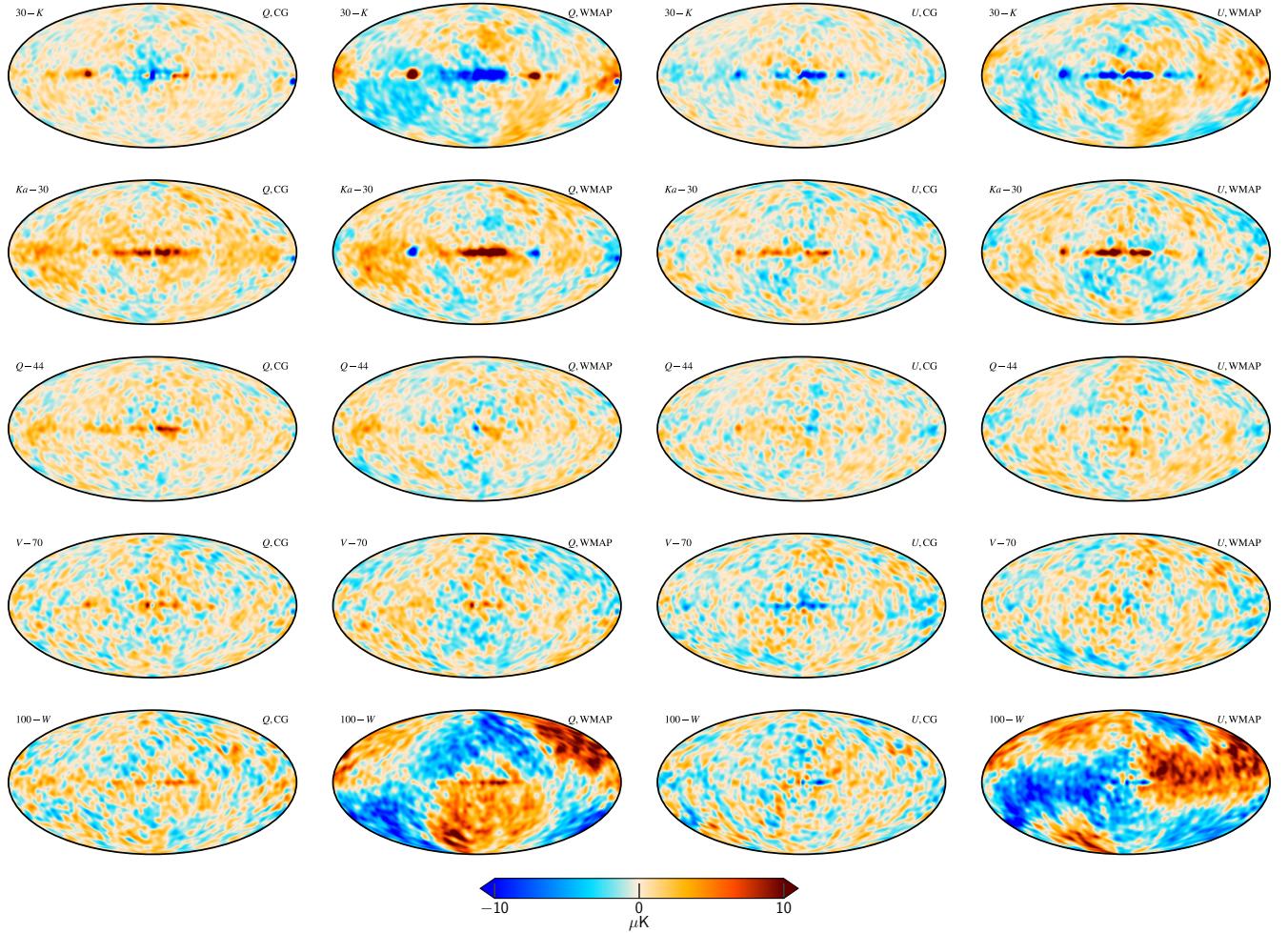


Fig. 23. Difference maps between similar WMAP and *Planck* frequency maps. The comparison plots go, by column: Stokes Q for the COSMOGLOBE sky maps, Stokes Q for official WMAP and BEYONDPLANCK data products, Stokes U for the COSMOGLOBEsky maps, and Stokes U for the official data products. (Top row) WMAP 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.

ples 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 6 and Fig. 26. We find that the dipole direction is consistent with the Colombo et al. (2022) result, but is $11 \mu\text{K}$ higher than the result from Hinshaw et al. (2009). Assuming a $3400 \mu\text{K}$ Solar 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 27 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 Λ CDM spectrum is plotted along side them. The middle panel shows the deviation from the *Planck* Λ CDM solution, in units of σ_ℓ from each individual pipeline, while the bottom panel shows the fractional difference with respect to the *Planck* Λ CDM spectrum.

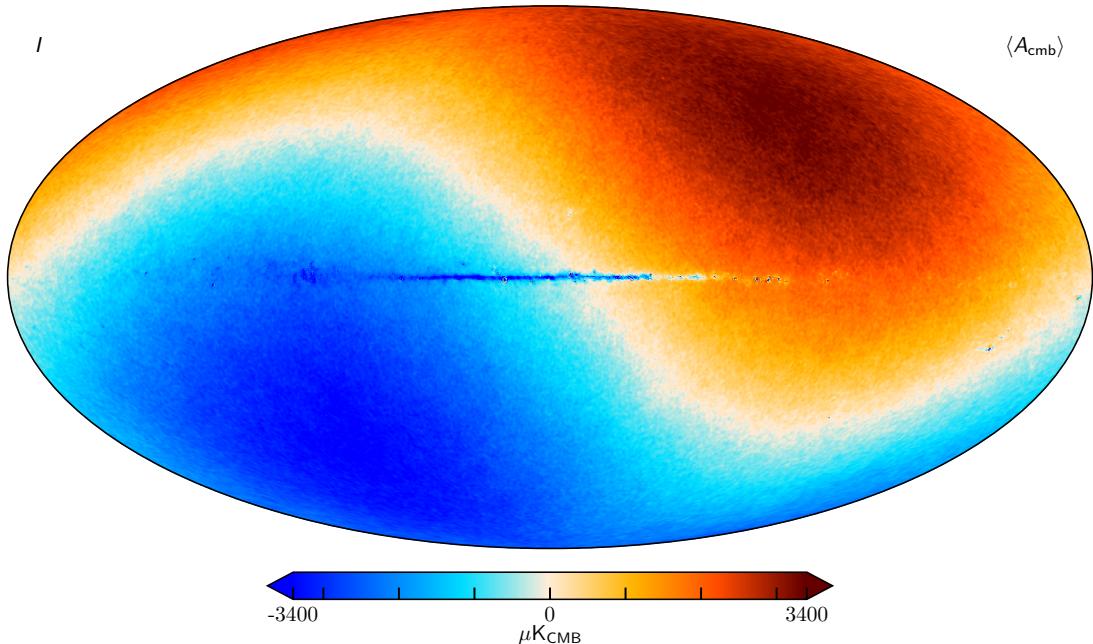


Fig. 24. Posterior mean CMB COSMOGLOBE temperature map, smoothed to an angular resolution of 14' FWHM.

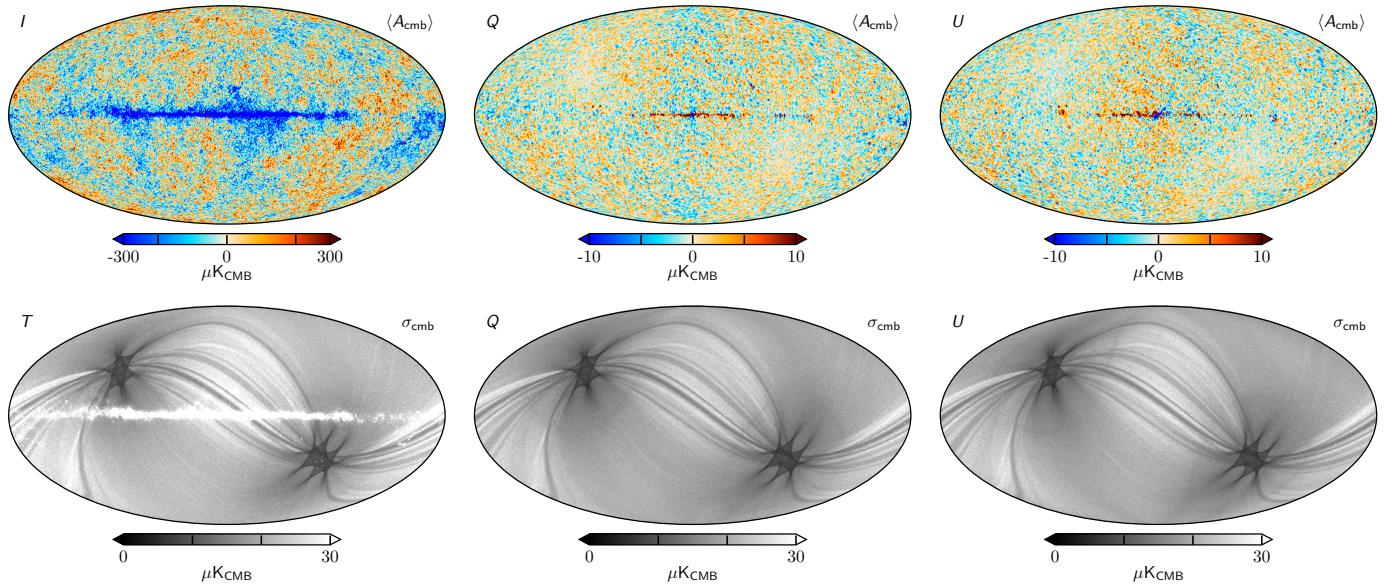


Fig. 25. Posterior mean CMB COSMOGLOBE maps and their standard deviation.

At $\ell \lesssim 500$, each of these datasets are signal-dominated, and each of the spectra agree at nearly every multipole. Notable 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 $\gtrsim 5\%$ higher than the ΛCDM 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- ℓ anomalies

Even though the CMB is well described by ΛCDM , there are several anomalies, especially at low multipoles, that seem to be in tension with ΛCDM . As a rule, the existence of these anomalies is not debated, but rather their significance within ΛCDM . Indeed, as argued by Bennett et al. (2011), the existence of these anomalies may simply be described using a posteriori predictions rather than confirming pre-existing predictions. The relevant question is not “How unlikely is ΛCDM given this effect’s existence?”, but rather, “How unlikely is this effect’s existence given ΛCDM ?”. Additionally, it is difficult in the standard methodology to determine how much of a given effect is due to unmodeled systematic errors. In this section, we use the full

Table 6. Comparison of Solar dipole measurements from *COBE*, *WMAP*, and *Planck*.

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

^a Statistical and systematic uncertainty estimates are added in quadrature.

^b Computed with a naive dipole estimator that does not account for higher-order CMB fluctuations.

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

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

^e 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.

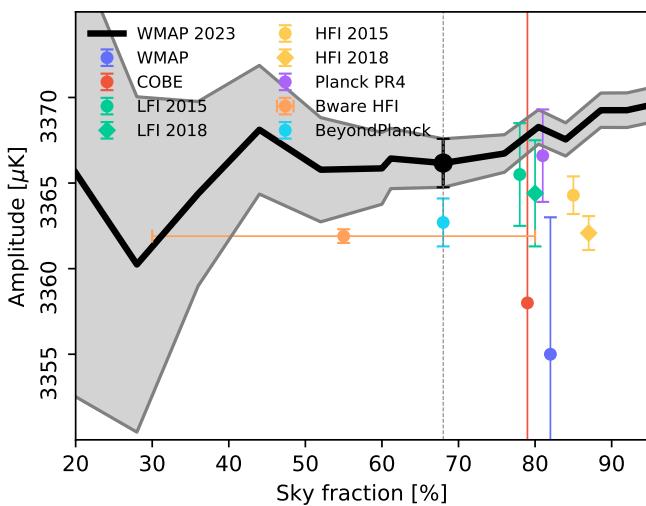


Fig. 26. CMB dipole amplitude as a function of sky fraction. The gray band indicates the 68 % posterior confidence region.

propagation of uncertainty to re-evaluate the significance of outstanding low- ℓ anomalies,

It has been noted since *COBE*-DMR that the quadrupole amplitude of our CMB is lower than expected from ΛCDM (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. Here we compare our estimate of the true realization of the quadrupole, σ_2 , with the mean value predicted by ΛCDM , $C_2^{\Lambda\text{CDM}}$. From the marginal distribution $P(\sigma_2 | \mathbf{d})$, we find that $\sigma_2 = 120 \pm 65 \mu\text{K}^2$, on its face a $\gtrsim 14\sigma$ deviation from the ΛCDM prediction of $C_2^{\Lambda\text{CDM}} = 1064.7 \mu\text{K}^2$. A more meaningful comparison is against the realizations of σ_2 observed given ΛCDM compared to the posterior probability distribution $P(\sigma_2 | \mathbf{d})$. To evaluate this, we draw 10^5 simulations of σ_2 given $C_2^{\Lambda\text{CDM}} = 1064.7 \mu\text{K}^2$. In Fig. 28, we show this comparison with

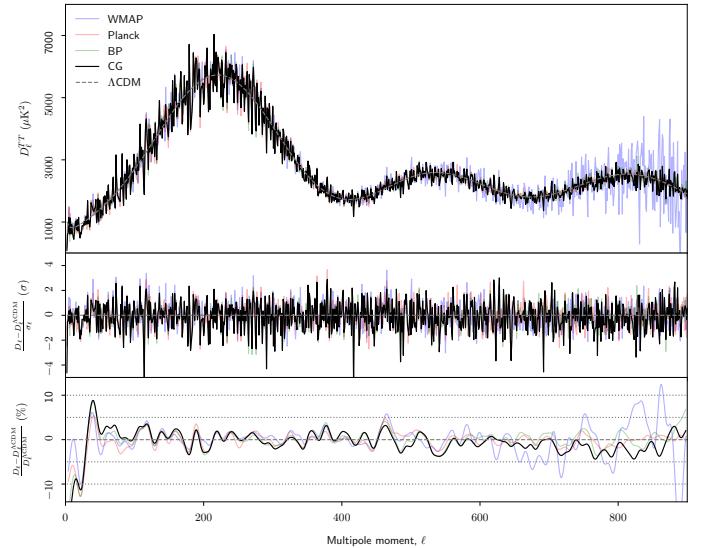


Fig. 27. (Top:) CMB temperature power spectrum, D_ℓ^{TT} , as derived by COSMOGLOBE (black), BEYONDPLANCK (green), *Planck* (red), and *WMAP* (blue). The best-fit *Planck* 2018 ΛCDM spectrum is shown in dashed gray. (Middle:) Residual power spectrum relative to ΛCDM , measured relative to the quoted error bars, $(D_\ell - D_\ell^{\Lambda\text{CDM}})/\sigma_\ell$. For pipelines that report asymmetric error bars, σ_ℓ is taken to be the average of the upper and lower error bar. (Bottom:) Fractional difference with respect to the *Planck* ΛCDM spectrum. In this panel, each curve has been boxcar averaged with a window of $\Delta\ell = 100$ to suppress random fluctuations.

the BEYONDPLANCK and COSMOGLOBE posterior distributions. Following Colombo et al. (2022), we compute the Blackwell-Rao likelihood for the true C_2 given the ensemble of observed σ_2 in Fig. 29. From this analysis, we find the probability-to-exceed (PTE) $C_2^{\Lambda\text{CDM}}$ to be 10 % for the COSMOGLOBE results. In comparison, the BEYONDPLANCK result had a PTE of 21 %.

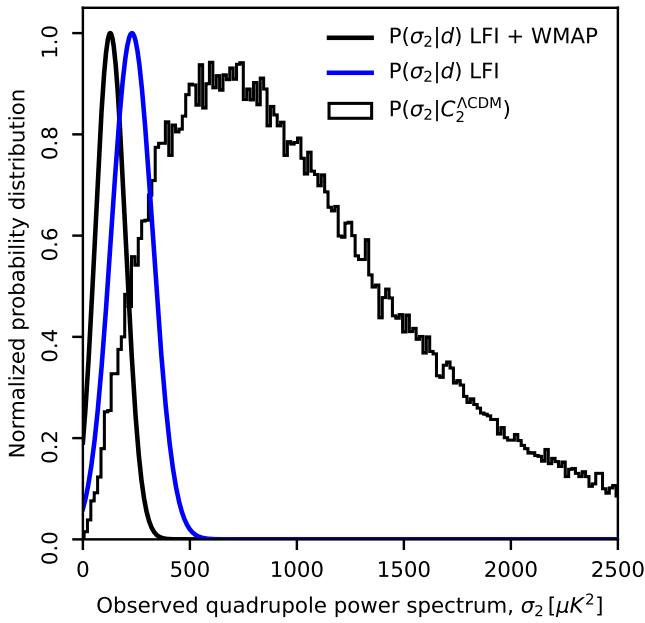


Fig. 28. Histogram of 100 000 realizations of C_2 given $C_2^{\Lambda\text{CDM}} = 1064.7$ compared with the measured power spectrum σ_2 of our universe for COSMOGLOBE (black) and BEYONDPLANCK (blue).

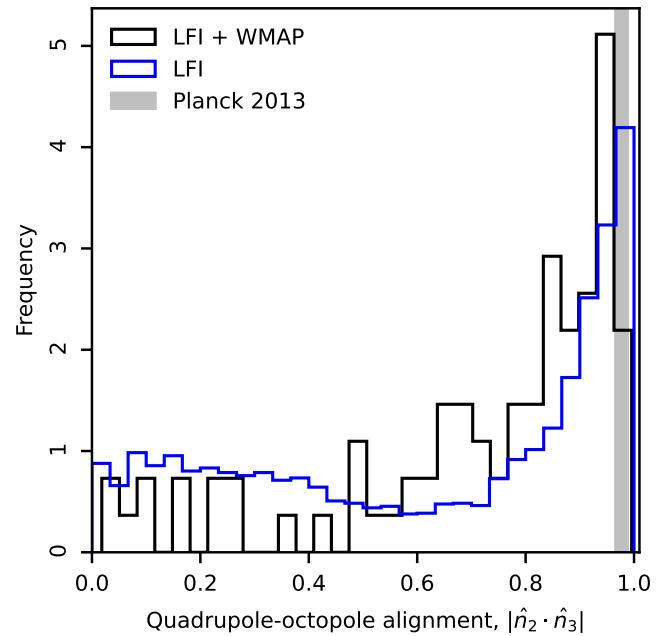


Fig. 30. The quadrupole-octopole alignment of COSMOGLOBE compared with BEYONDPLANCK.

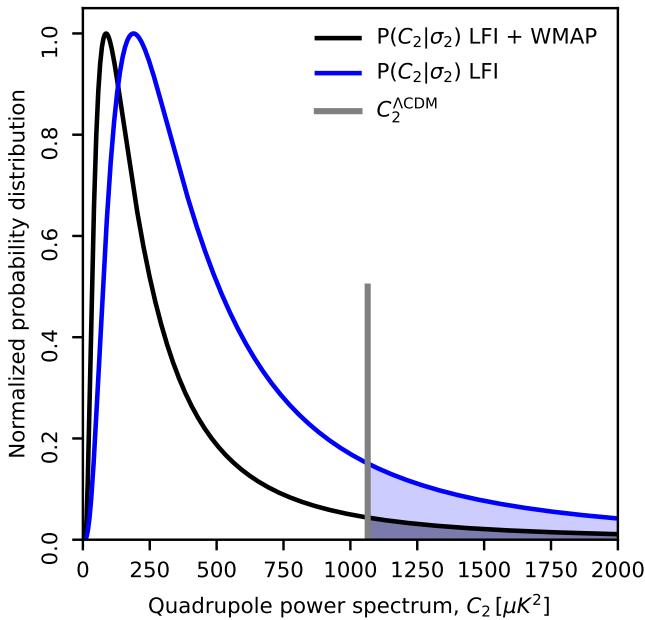


Fig. 29. Marginal probability distribution of the ensemble-averaged C_2 given the data, $P(C_2|d)$, as measured by COSMOGLOBE (black) and BEYONDPLANCK (blue).

6.2. Galactic foregrounds

As described in Sect. 2.5, 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. The thermal dust amplitude is also fit here, though the addition of *Planck* 857 GHz provides most of the constraining power. The results for each of these foreground components are displayed in Fig. 33. Though the form of the AME model

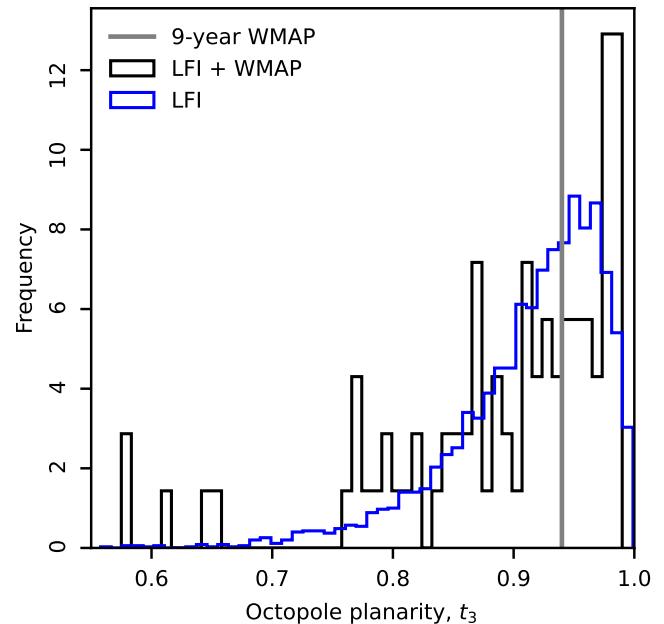


Fig. 31. The octopole planarity statistics t_3 compared with the BEYONDPLANCK analysis (blue).

is different in this work, the morphologies of the components match well with those presented in Andersen et al. (2022). This is an unsurprising result as the only true difference other than the AME model is the addition of the *WMAP* K -band to this analysis.

In polarization, the addition of *Planck* 353 GHz provides an anchor for the polarized thermal dust emission, and as a result only the polarized synchrotron emission amplitude is explicitly fit in this work. The Stokes Q and U synchrotron amplitude mean and standard deviations are shown in Fig. 34. The mean amplitude map is in good agreement with that presented in Svalheim

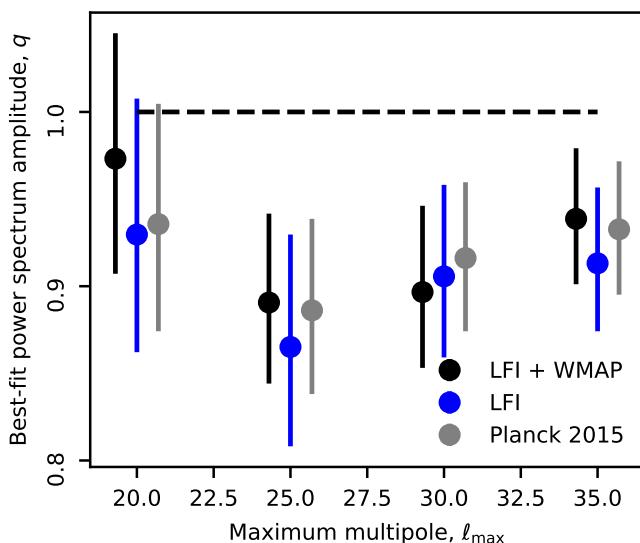


Fig. 32. Best-fit amplitude, q , of the low multipole power spectrum $C_\ell = qC_\ell^{\Lambda\text{CDM}}$, $2 \leq \ell \leq \ell_{\max}$ compared to *Planck* 2015 (grey) and BEYONDPLANCK (blue).

et al. (2022b), though the morphology of the standard deviation map matches more closely with that presented in Herman et al. (2022b), which shows a combination 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.

We can assess the quality of the component separation procedure is evaluated through a reduced- χ^2 map, shown in Fig. 35, as well as through the map space residuals Fig. B.6. The χ^2 per pixel is plotted in the form $(\chi^2 - n_{\text{dof}})/\sqrt{2n_{\text{dof}}}$, where n_{dof} is approximately the number of degrees of freedom, including the number of full-frequency pixels within each $N_{\text{side}} = 128$ pixel, minus the number of fitted parameters. As noted by, e.g., Planck Collaboration X (2016) and Svalheim et al. (2022b), the total n_{dof} is difficult to rigorously compute, due spatial correlations and priors induced. To aid in interpretation, we find $n_{\text{dof}} = 900$ is a good approximate fit to the χ^2 outside of the K-band processing mask. For comparison, the total number of subpixels within each $N_{\text{side}} = 128$ superpixel is approximately 976, which is in fair agreement with 900 when taking into account smoothing of each beam.

Overall, we find that the model appears well at high Galactic latitudes high-significance deviation associated with radio point sources. In addition, there is a clear excess along the Galactic plane. As the Galactic Center is difficult to model correctly, it is not surprising that the parametric fit used in this paper failed to characterize low-latitude emission. Overall, we find that the goodness-of-fit failures are not associated with instrumental effects. In particular, there is no trace of the poorly measured modes in the reduced- χ^2 map, indicating that they have been properly mitigated and characterized by the RMS estimation.

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

With a robust estimate of the sky model within the instrumental uncertainties, we can quantitatively assess the relative importance of each channel for constraining individual foreground components. In Fig. 36, we **how do we do this exactly?** to find the relative constraining power each individual band from K to

W inclusive contributes to the foreground fit. In particular, by comparing the relative uncertainty of each band to the amplitude of the signal we expect it to observe, we can assign a ranked relative importance of each band within the component separation solution. Note that we leave out Haslam and *Planck* 353 and 857 GHz in this comparison, as they essentially act as pure noiseless templates in this framework.

For the CMB and thermal dust, we find that LFI's 70 GHz channel contains the most constraining power, both in temperature and polarization. In particular, thermal dust is constrained better by 70 GHz than W-band due to its lower white noise level. More surprising is the strength of K-band to constrain all low-frequency foregrounds in both intensity and polarization. At 23 GHz, the K-band measurement is the lowest-frequency full-sky map of polarized synchrotron emission, which has been rightly recognized as part of the longstanding legacy of the *WMAP* experiment. However, the strength of K-band with respect to, e.g., the LFI 30 GHz channel in intensity is more surprising, as the white noise level of 30 GHz is lower than that of K-band, and the spectral dependence of free-free and AME is less dramatic than for synchrotron.

Say a bit more about the long-lasting legacy of K-band

7. Systematic error corrections and uncertainties

7.1. Sky map corrections

7.2. Power spectrum residuals

8. Outstanding issues

8.1. Noise modeling

8.2. V-band quadrupole residual

8.3. Degeneracy between K-band calibration and AME dipole

8.4. Other minor effects

8.4.1. Time-variable bandpass modeling

8.4.2. Polarized sidelobe modeling

9. 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]

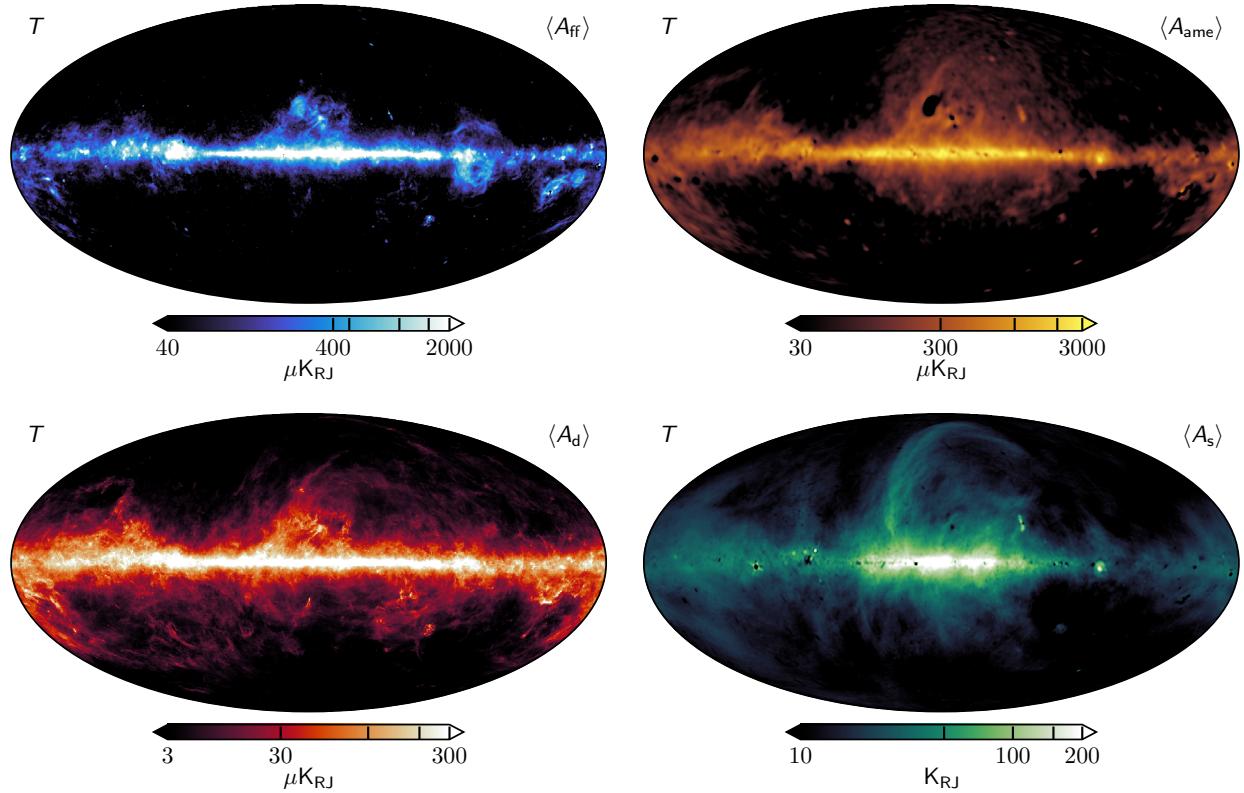


Fig. 33. Foreground intensity maps, evaluated at their respective reference frequencies. (Top left) Free-free emission at 40 GHz. (Top right) Anomalous microwave emission evaluated at 22 GHz. (Bottom left) Thermal dust emission at 70 GHz. (Bottom right) Synchrotron emission evaluated at 408 MHz.

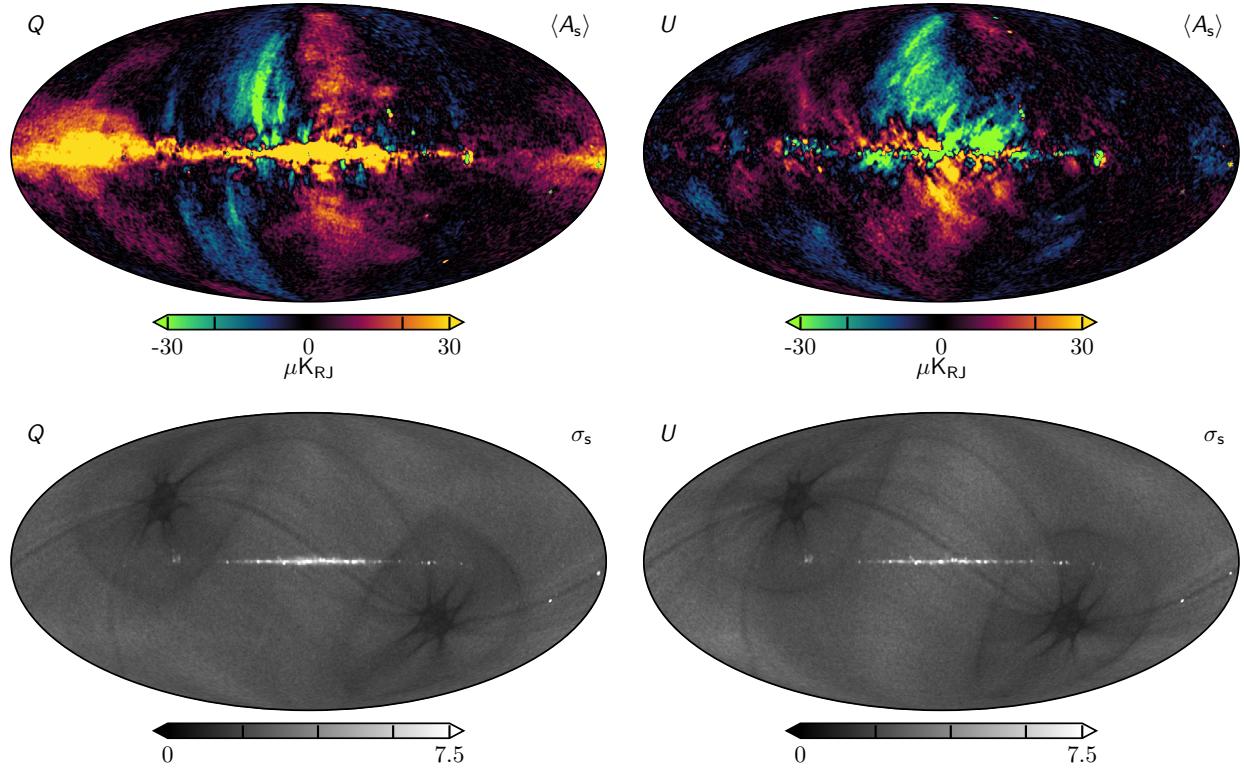


Fig. 34. Polarized synchrotron maps and their standard deviations evaluated at 30 GHz.

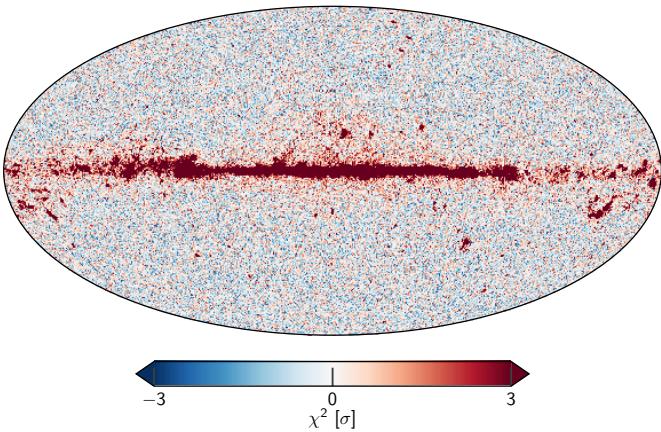


Fig. 35. Reduced- χ^2 , using $n_{\text{dof}} = 900$, which comes from fitting to the regions outside of the K -band processing mask.

- Gjerløw 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. 2022a, A&A, in press [arXiv:2203.13635]
 Herman et al. 2022b, 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., 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)
 Rubiño-Martín, J. A., Guidi, F., Génova-Santos, R. T., et al. 2023, MNRAS, 519, 3383
 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-Haimoud, 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:2201.08503]
 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

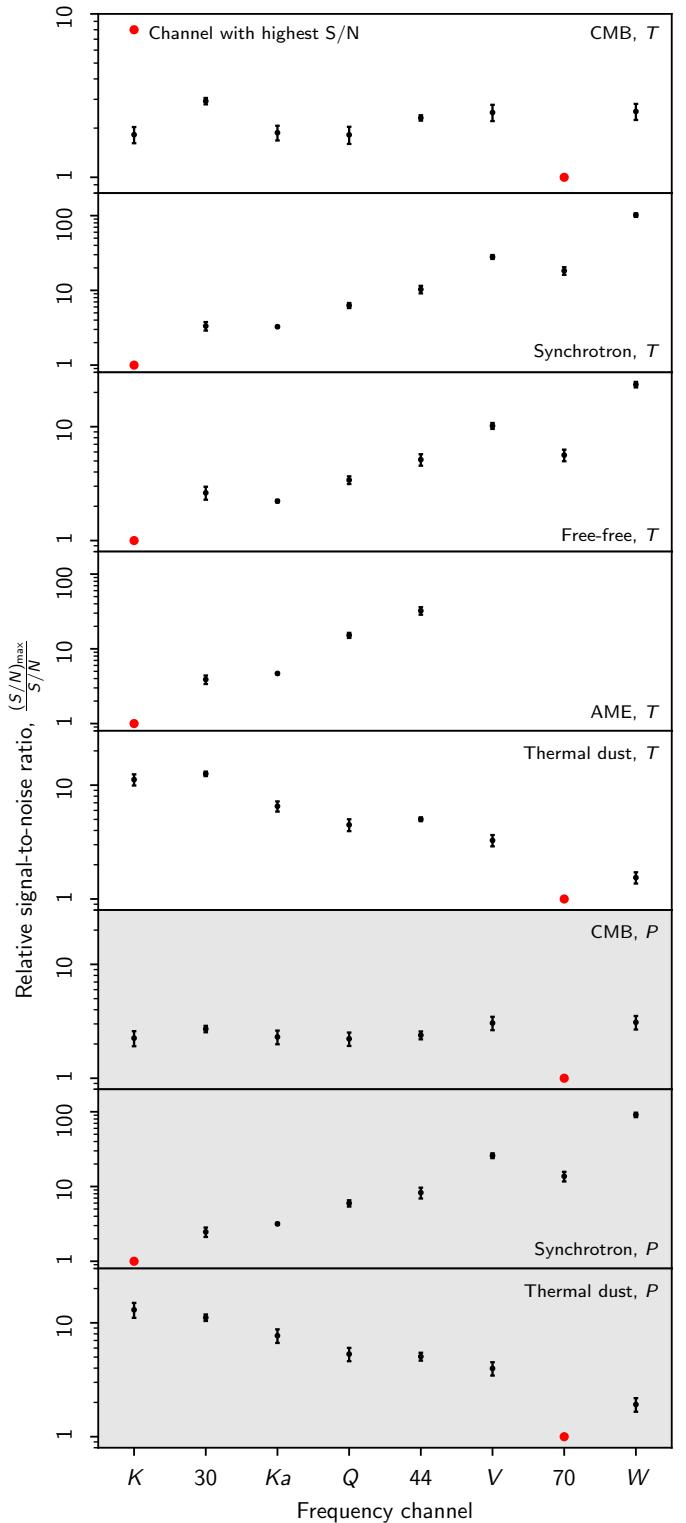


Fig. 36. Relative signal-to-noise ratios for WMAP and LFI channels and various components.

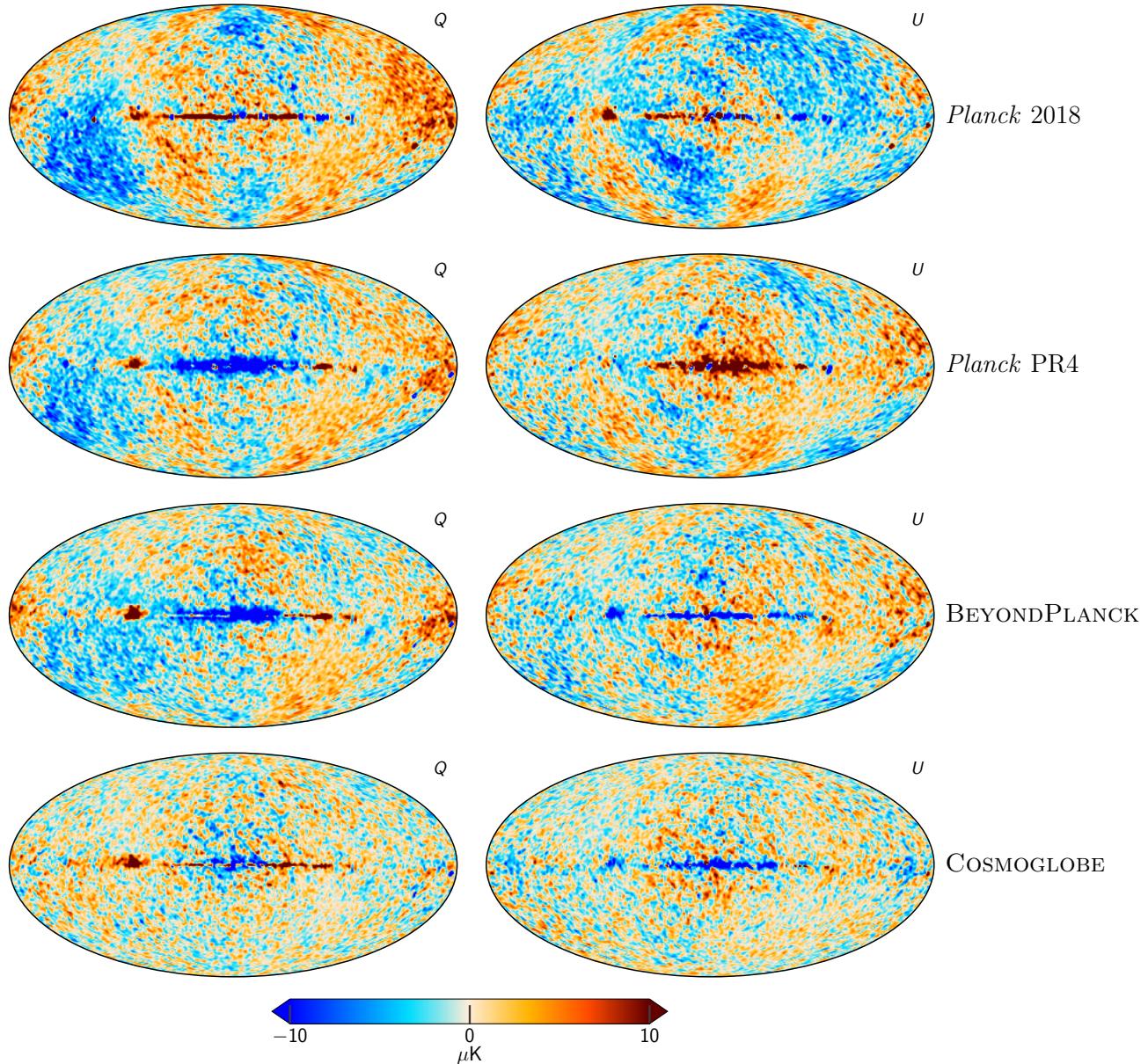


Fig. 37. Difference maps between the *Planck* 30 GHz and *WMAP* K-band maps. The columns are (1) *Planck* 2018 v. *WMAP*9, (2) *Planck* PR4 v. *WMAP*9, (3) BEYONDPLANCK v. *WMAP*9, and (4) COSMOGLOBE *Planck* 30 GHz and *WMAP* K-band both produced in this paper. All maps have been smoothed to a common resolution of 2° 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 χ^2

Appendix A.2: Transmission imbalance

Appendix B: *WMAP* frequency map survey

Appendix C: Comparison with BEYONDPLANCK LFI results

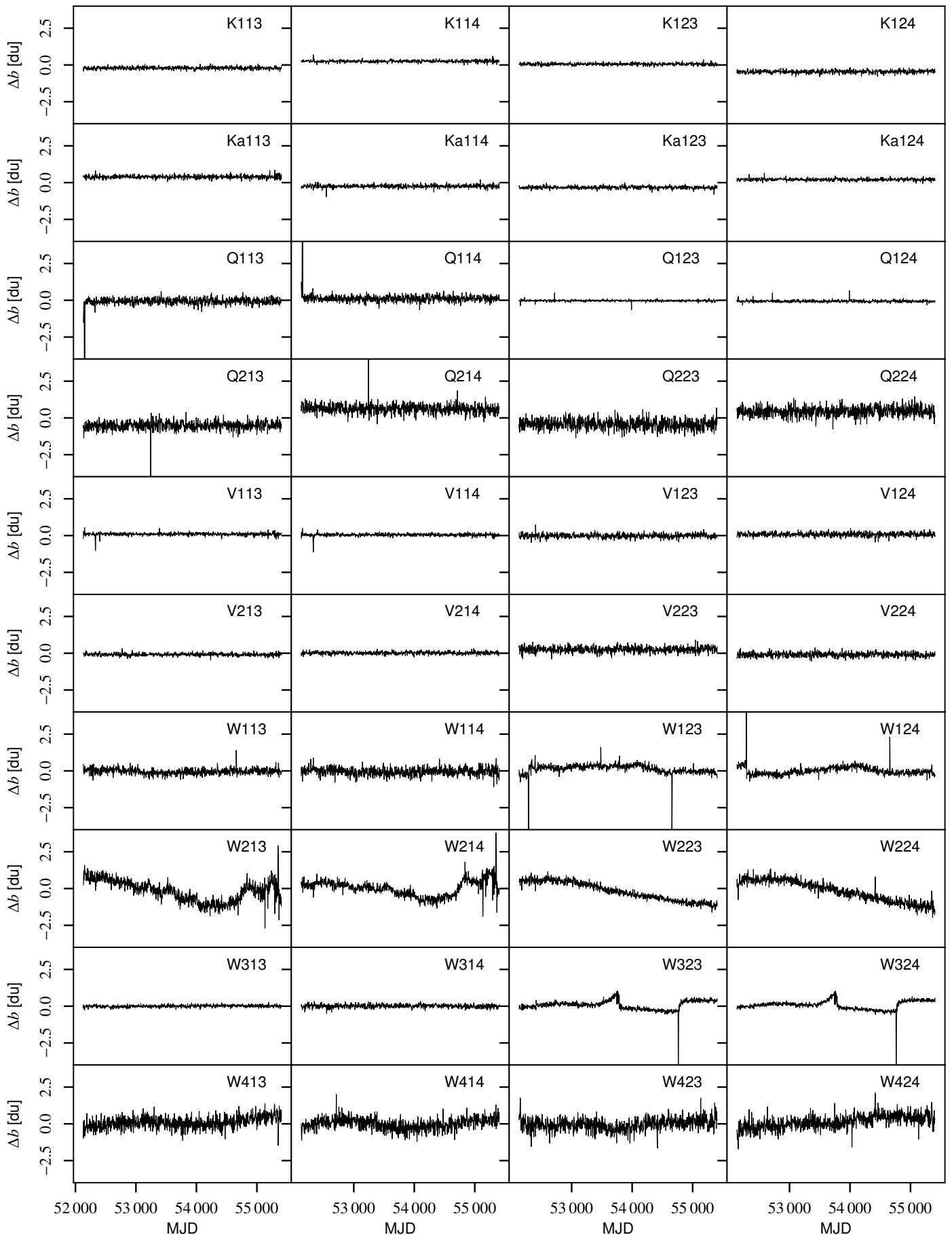
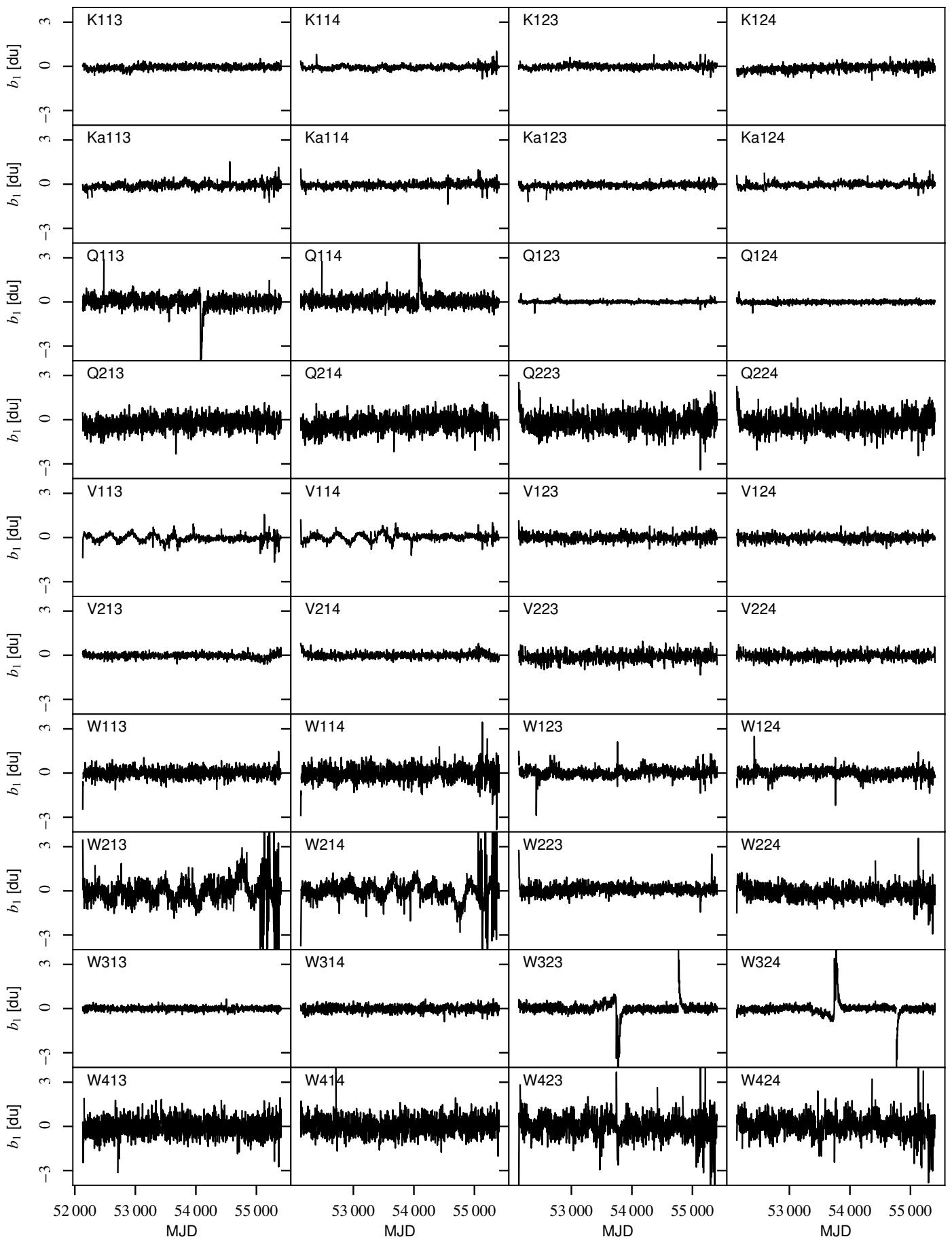
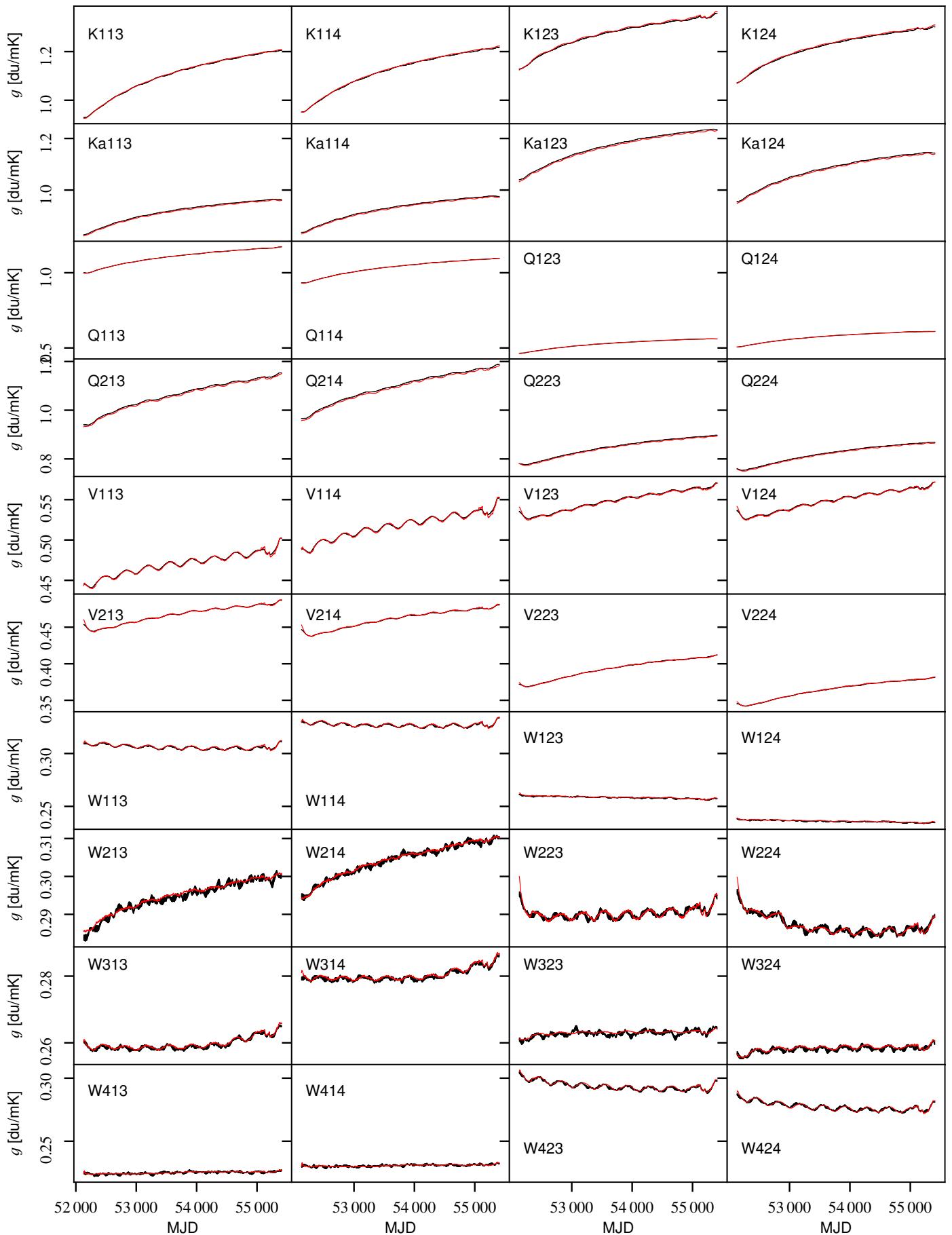


Fig. A.1. Difference in baseline solution, $b_0^{\text{CG}} - b_0^{\text{WMAP}}$.

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

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

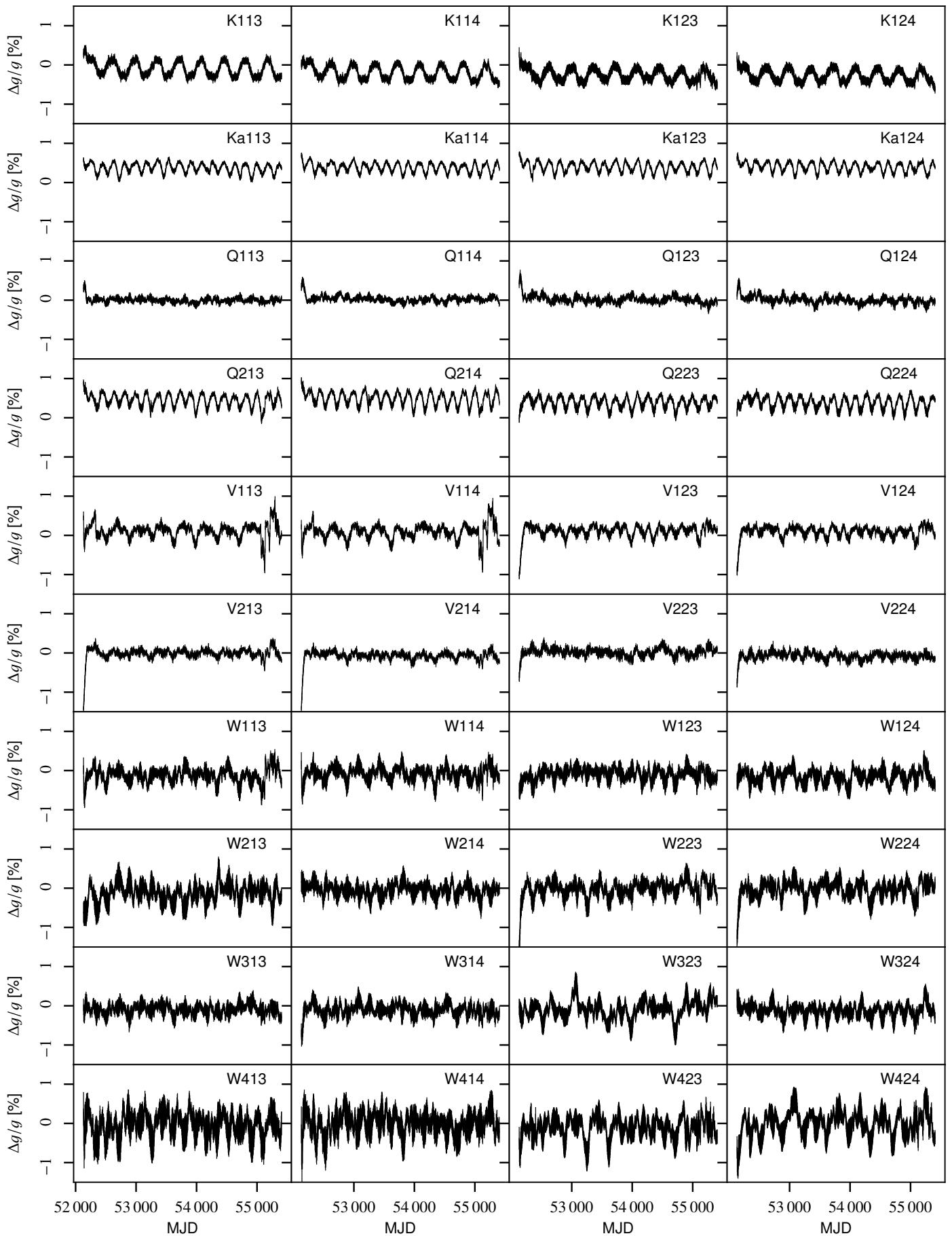
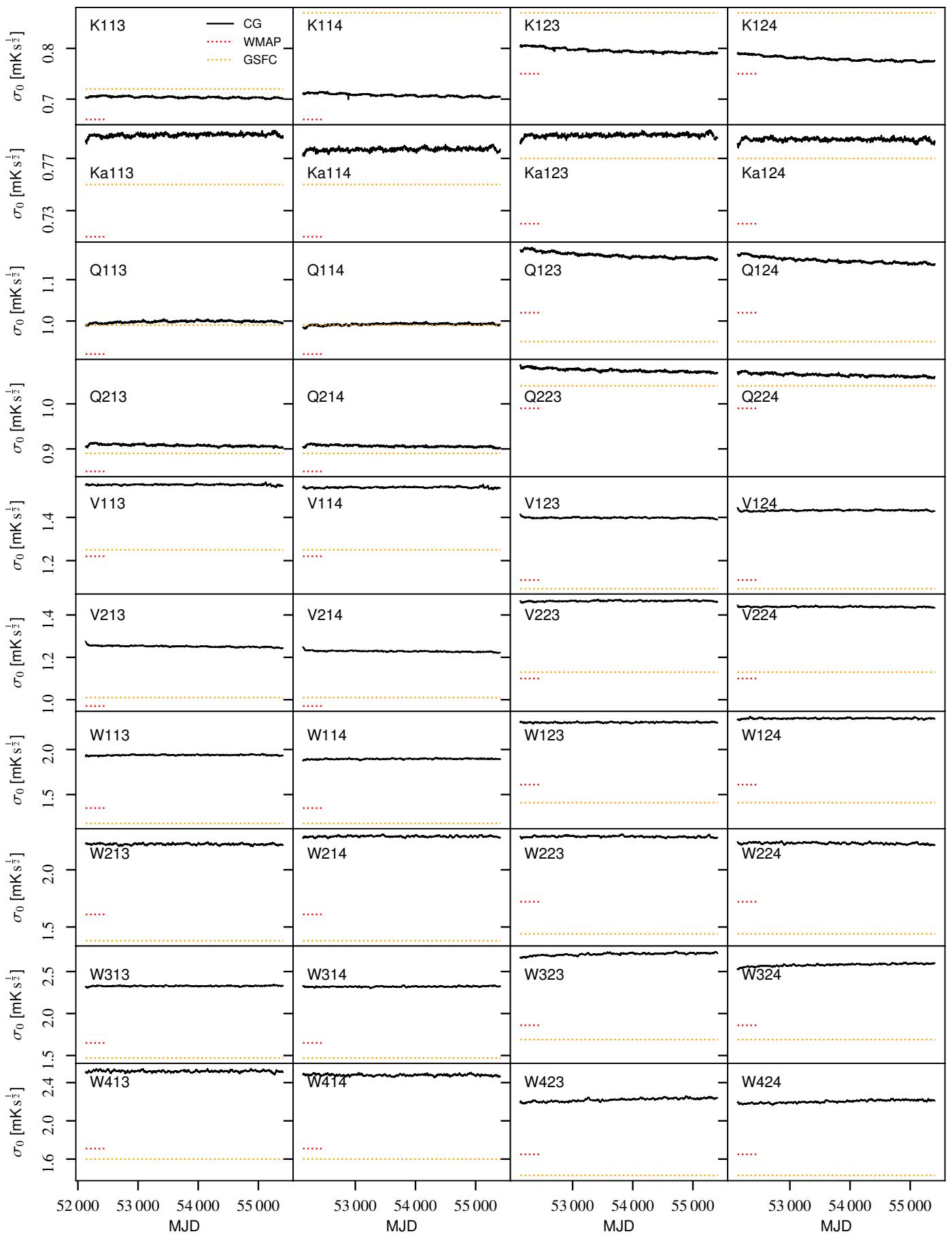
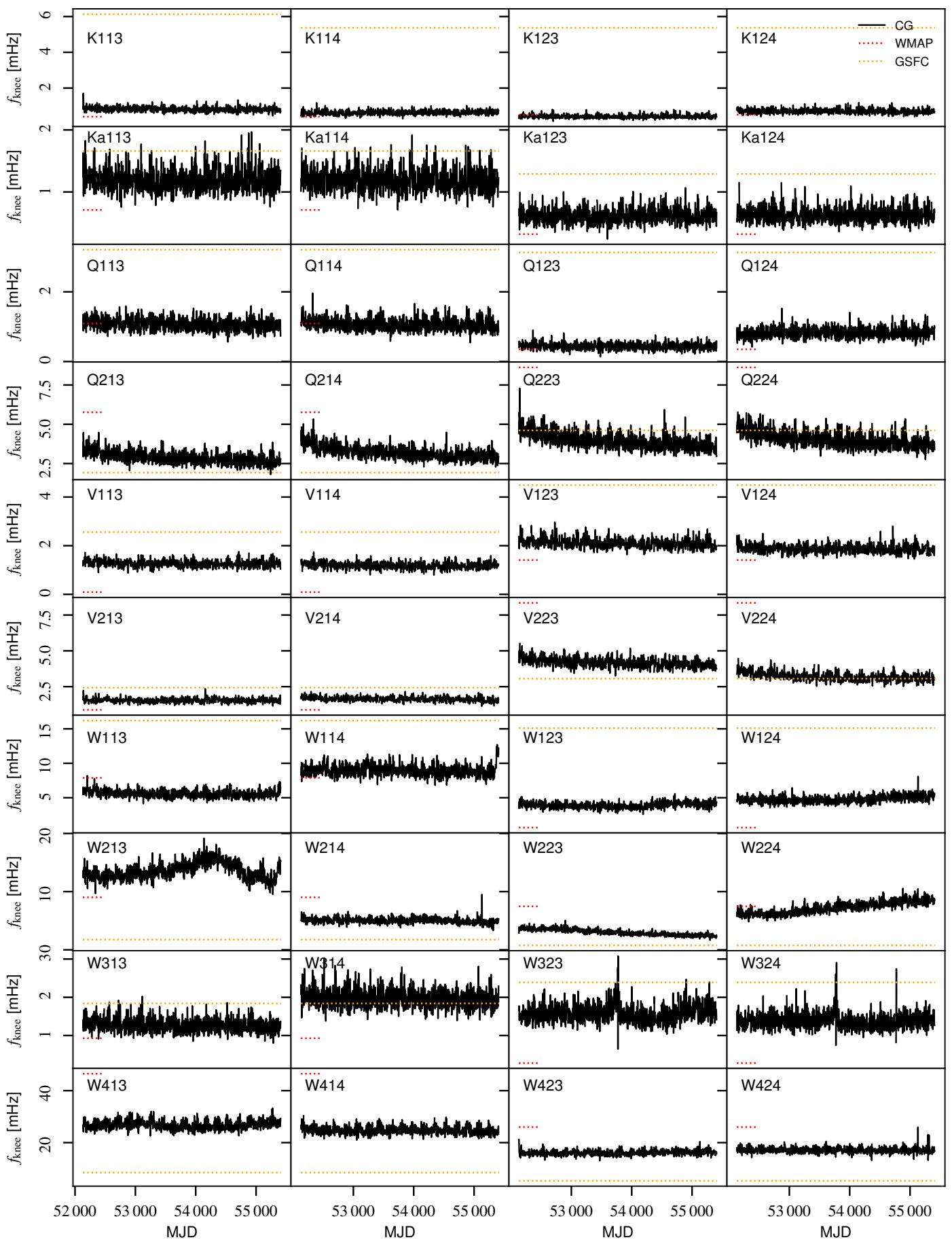
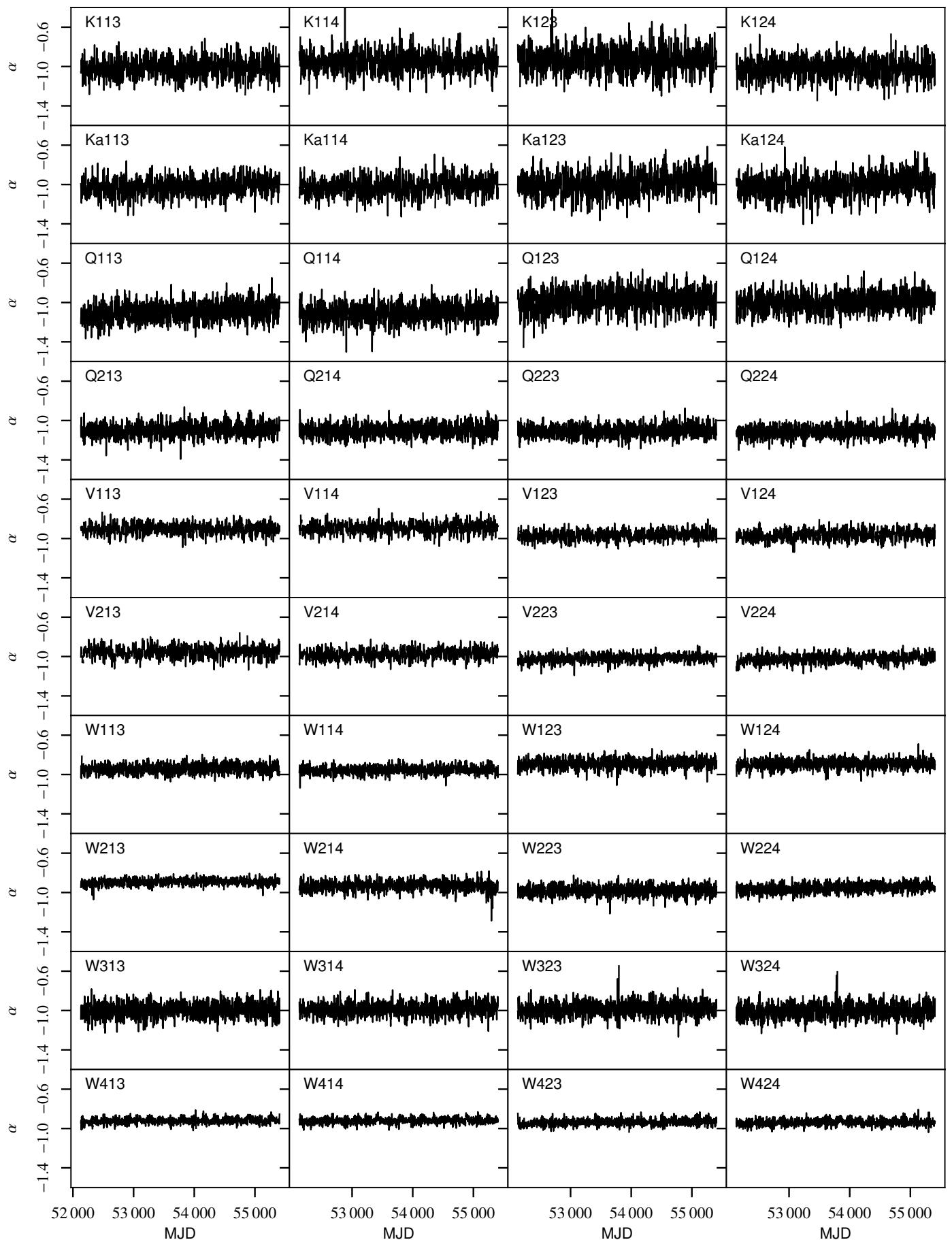
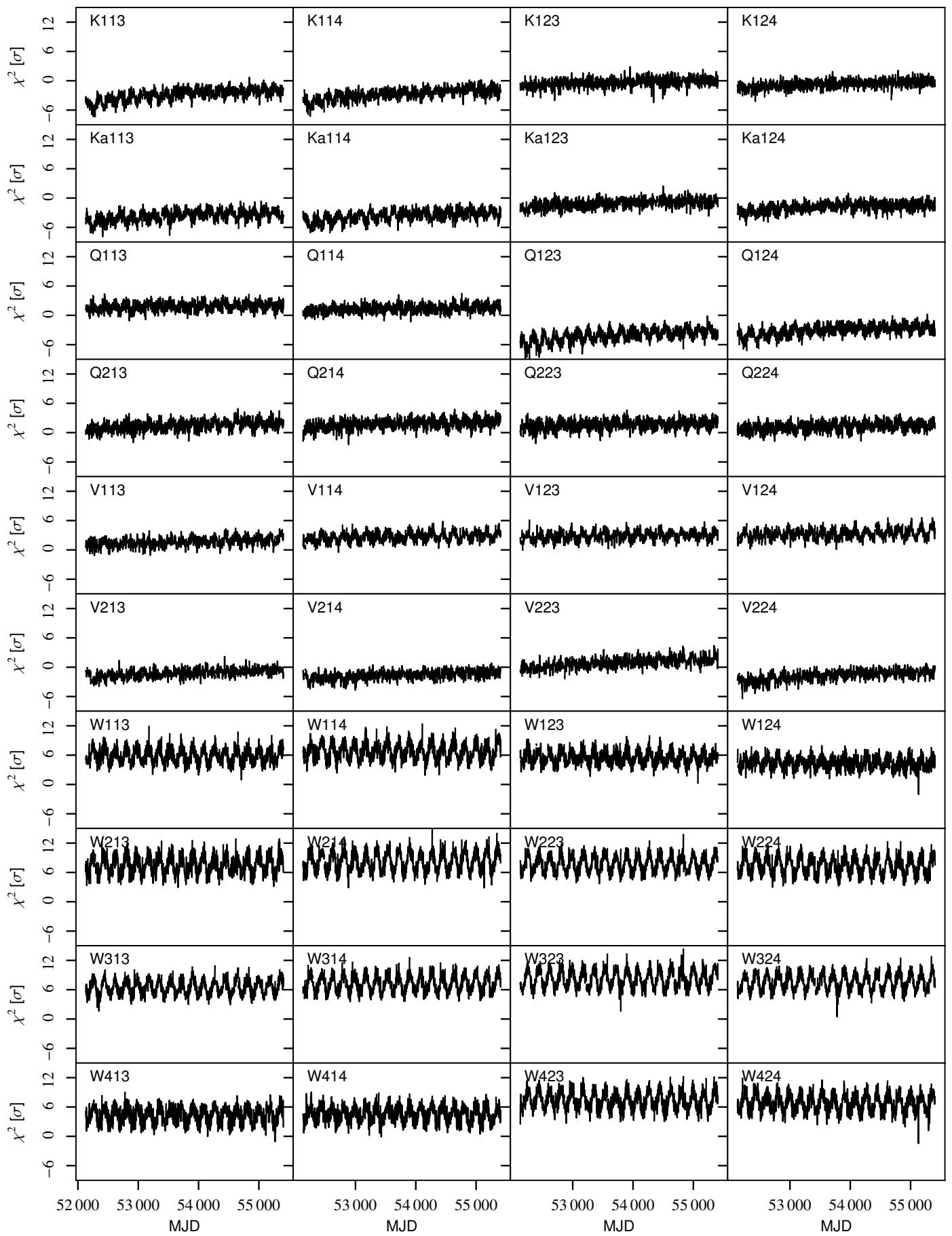


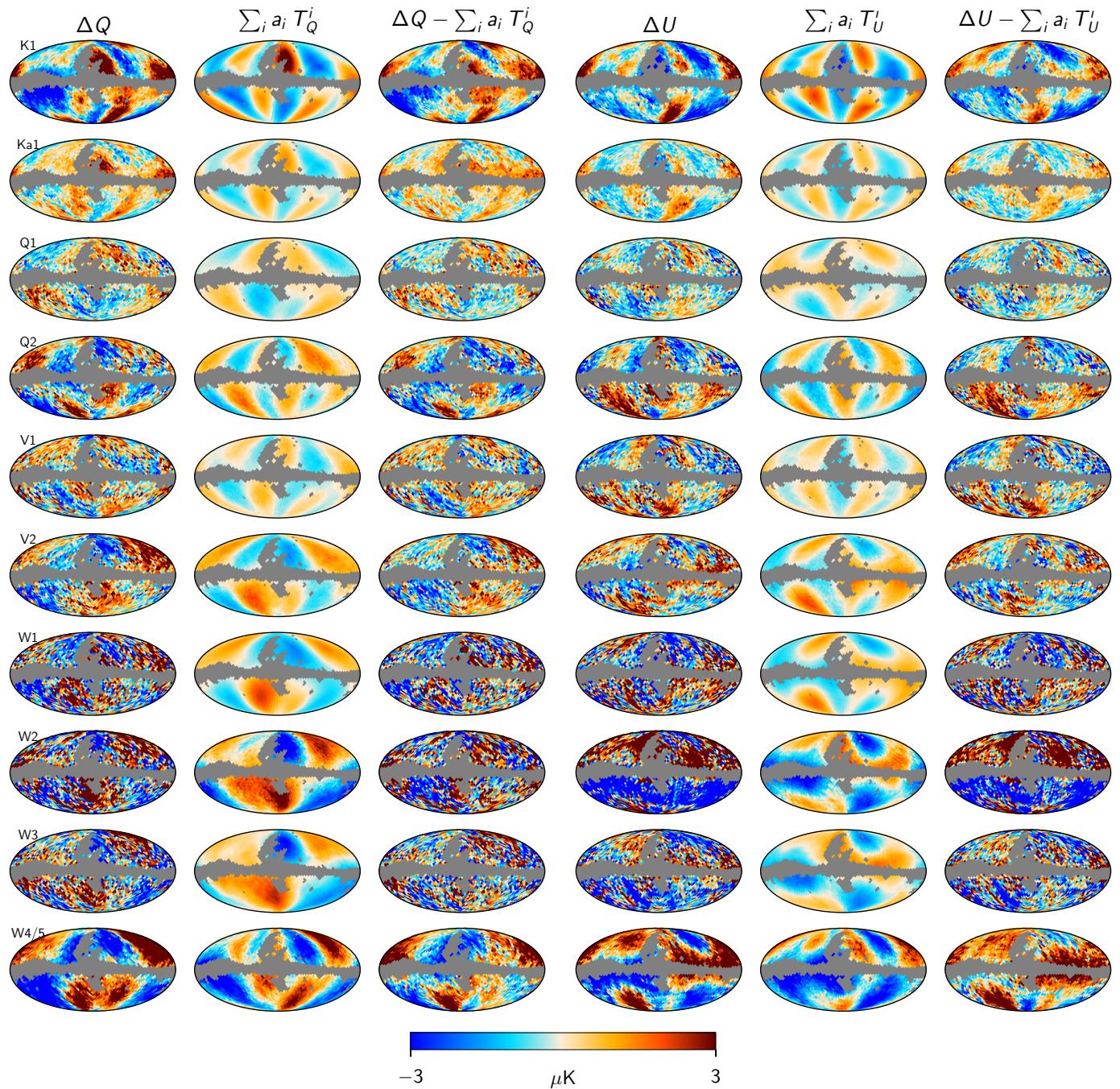
Fig. A.4. Relative difference in gain solutions, $(g^{\text{CG}} - g^{\text{WMAP}})/g^{\text{WMAP}}$.

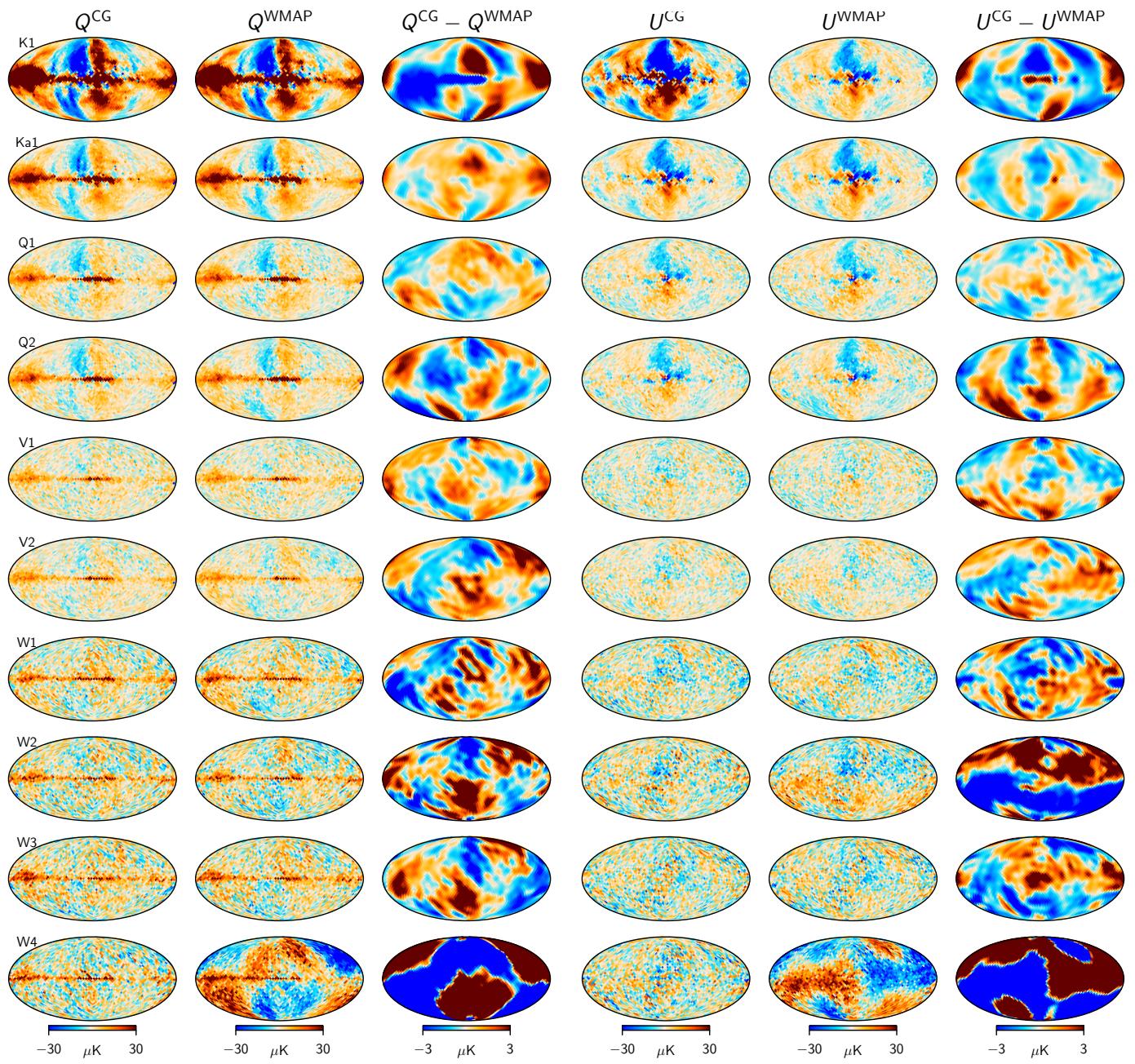
**Fig. A.5.** σ_0 .

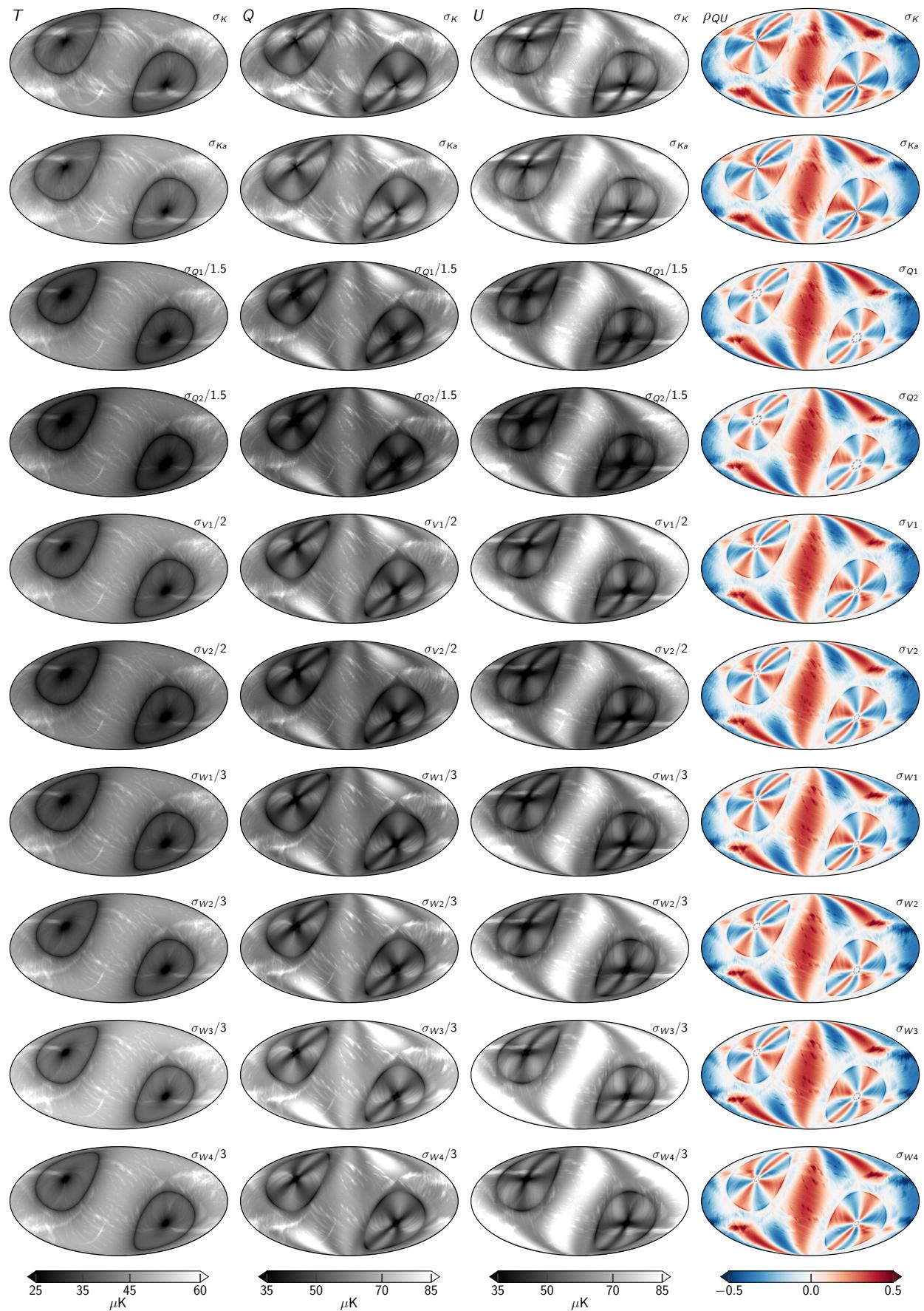
**Fig. A.6.** Fknee.

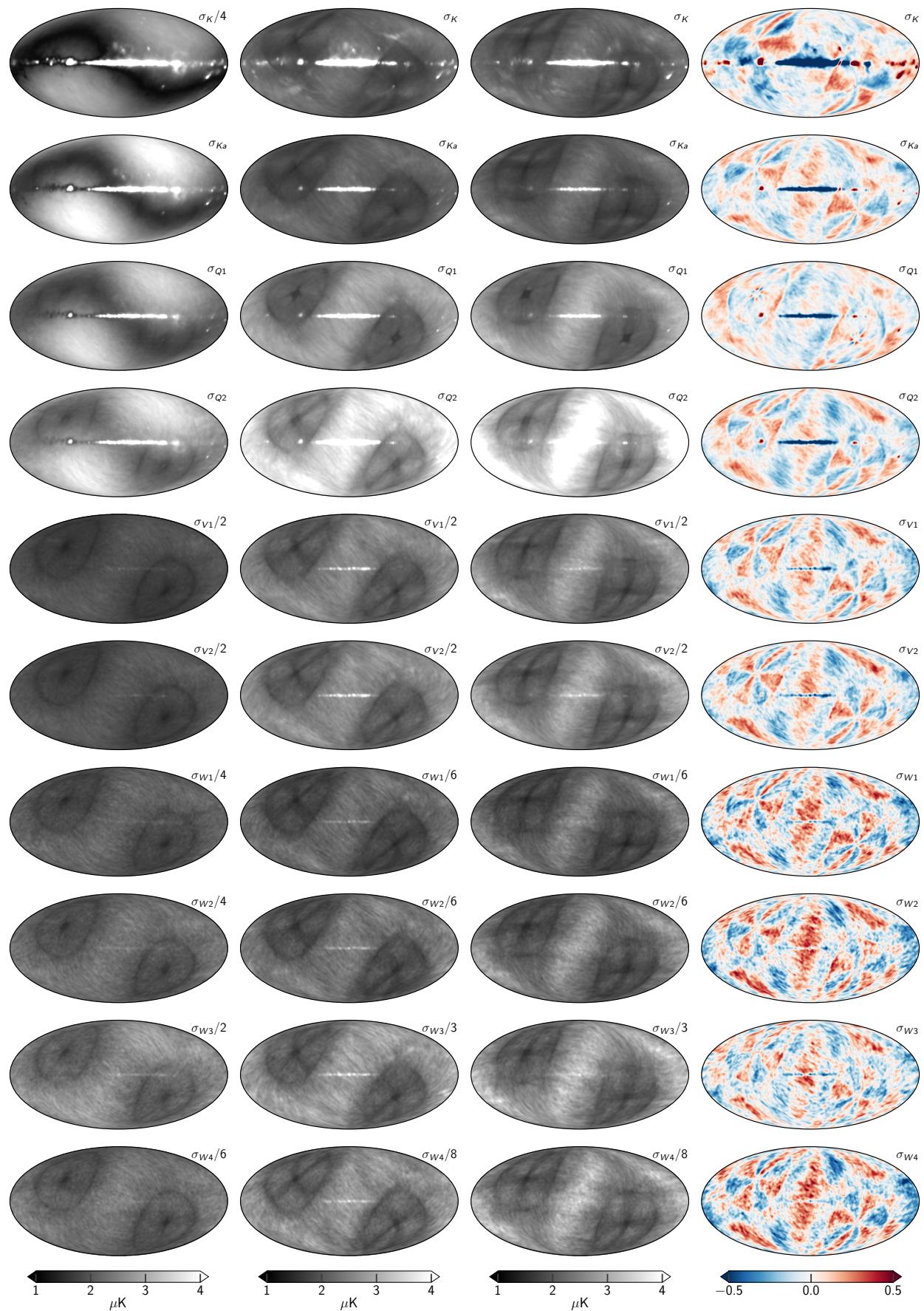
**Fig. A.7.** alpha.

**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

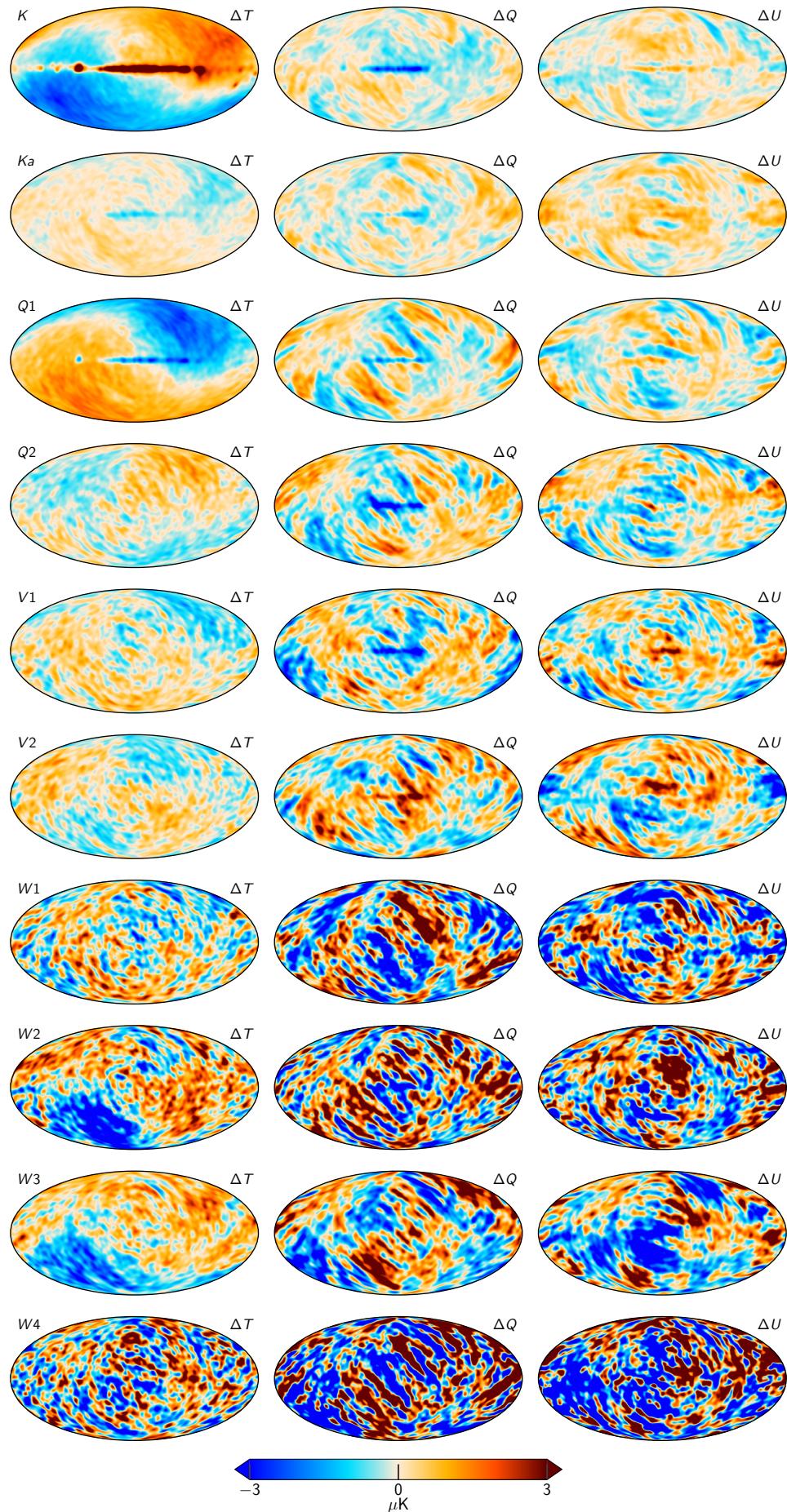
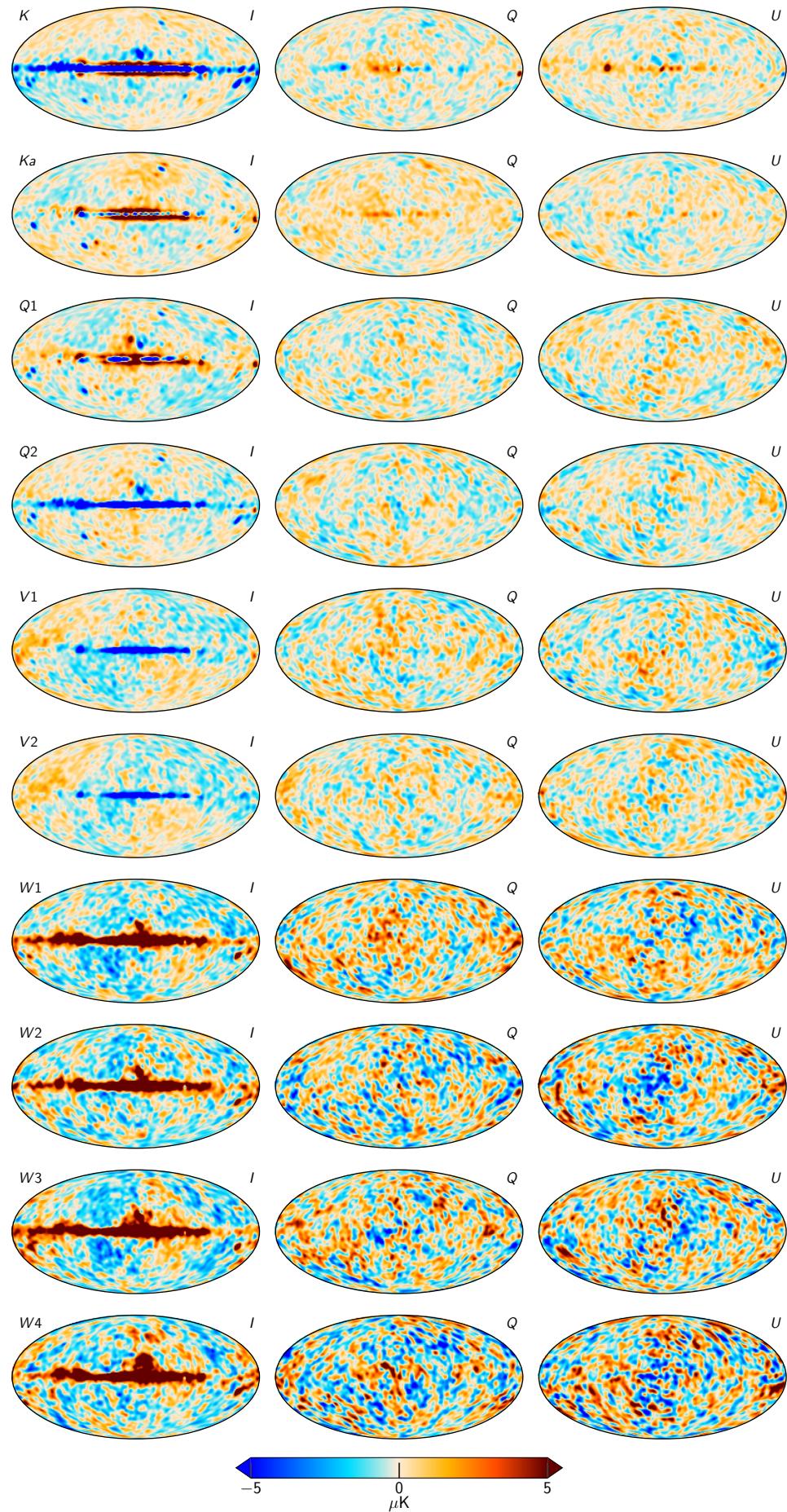


Fig. B.4. Differences between two samples
Article number, page 52 of 54

**Fig. B.5.** TOD Residuals for each of the WMAP channels, smoothed by 5° .

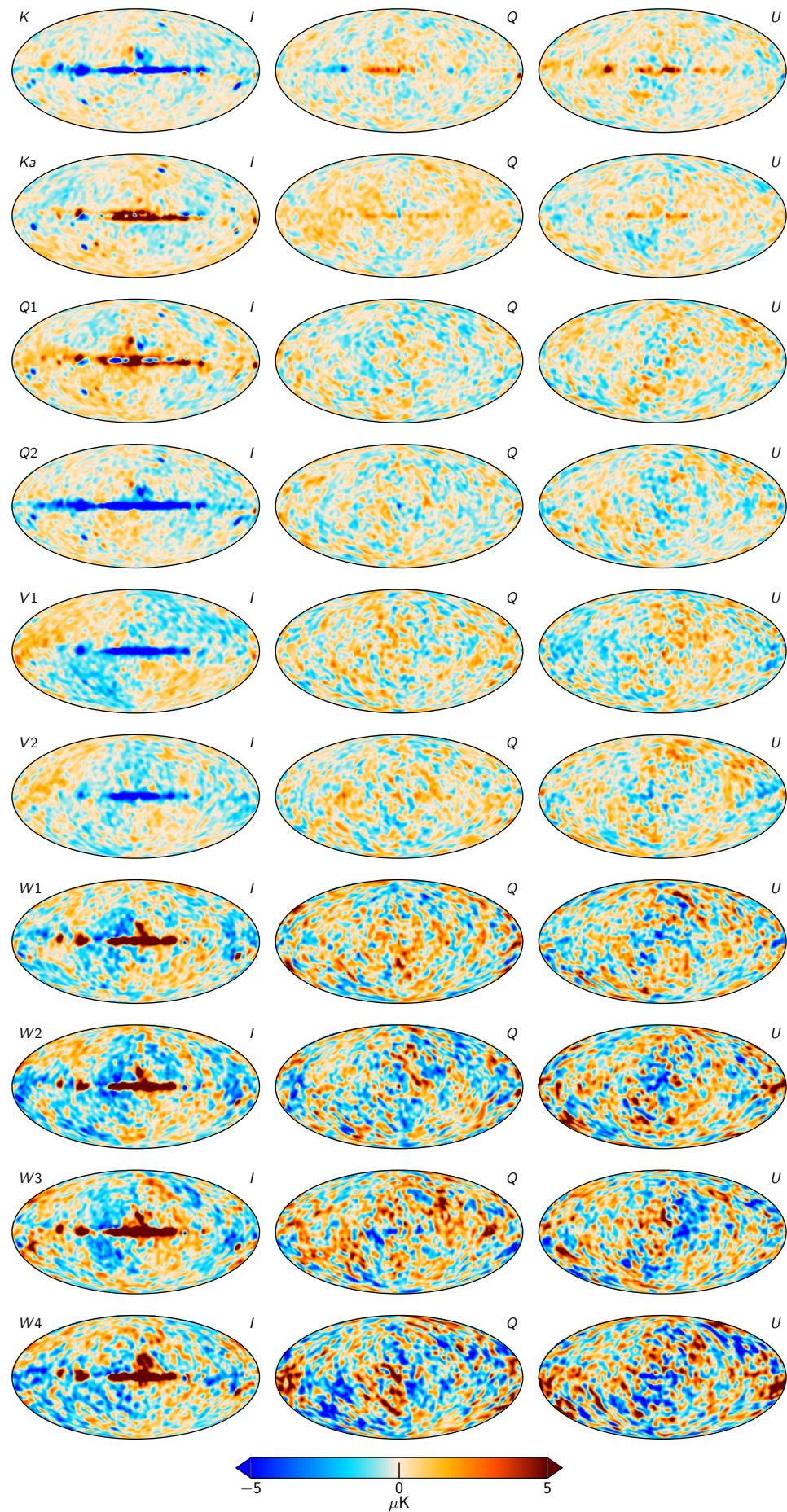


Fig. B.6. Component separation residuals for each of the WMAP channels, smoothed by 5° .
Article number, page 54 of 54