

Sparse estimation of model-based diffuse thermal dust emission

Melis O. Irfan,¹★ Jérôme Bobin,¹

¹*Laboratoire CosmoStat, AIM, UMR CEA-CNRS-Paris 7 Irfu, SAP/SEDI, Service d'Astrophysique, CEA Saclay, F-91191 GIF-SUR-YVETTE CEDEX, France.*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Component separation for the *Planck* HFI data is primarily concerned with the estimation of thermal dust emission, which requires the separation of thermal dust from the cosmic infrared background (CIB). Current estimation methods make use of smoothing, through convolution with a Gaussian beam, to decouple thermal dust emission from CIB anisotropies at small angular scales. In this paper we present a new parameter estimation method, *premise*: Parameter Recovery Exploiting Model Informed Sparse Estimates. This method exploits the sparse nature of thermal dust emission to calculate all-sky maps of thermal dust temperature, spectral index and optical depth at 353 GHz. The data used are simulation data so as to validate *premise*. We find the percentage difference between the *premise* results and the true values to be 2.8, 5.7 and 7.2 per cent at the 1σ level across the full sky for thermal dust temperature, spectral index and optical depth at 353 GHz, respectively. Comparison between *premise* and a GNILC-like method over select regions of the sky reveal *premise* to outperform with increasing success as the signal-to-noise ratio worsens.

Key words: Cosmology: diffuse radiation, ISM: dust, extinction, Methods: Data Analysis, Methods: Statistical

1 INTRODUCTION

Within our Galaxy dust grains, such as carbonaceous and amorphous silicates (Draine & Li 2007), are heated by inter-stellar, UV radiation. The resulting emission, known as thermal dust emission, is the dominant diffuse Galactic emission at frequencies > 100 GHz (Bennett et al. 2013). Measurements of thermal dust emission reveal the chemistry of the interstellar medium (Compiègne et al. (2011); Jones et al. (2013)), trace interstellar radiation and reveal dust mass and dust column density. Additionally, thermal dust emission is a prominent astrophysical foreground present in measurements of the cosmic microwave background (CMB). In order to infer cosmological information from measurements of the CMB over 100 GHz, accurate subtraction of thermal dust emission is crucial (Planck Collaboration et al. 2016a).

Over large angular-scales thermal dust emission can be modelled as blackbody emission, as can the cosmic infrared background (CIB); diffuse infrared emission from dusty galaxies across all redshifts. The CIB is a strong tracer of the star formation history of the Universe (Lagache et al. 2005) and its anisotropies shed light on galaxy clustering and dark matter halo distributions (Béthermin et al. 2013). The similarity in spectral shape between thermal dust emission and the CIB adds complexity to their separation. The correlation between thermal dust and H_I emission can be used to differentiate between thermal dust and the CIB at large angular scales

(Planck Collaboration et al. 2014) but unfortunately this correlation cannot be exploited at scales smaller than several degrees.

Since the second *Planck* data release specific attempts to separate thermal dust emission from the CIB over *Planck* HFI frequencies have been implemented with increasing success. Planck Collaboration et al. (2014) fit a modified blackbody (MBB) to the data and present all-sky maps of the three fit parameters: dust temperature, spectral index and optical depth at 353 GHz. The MBB parameters do not represent the mass-weighted averages of the physical dust properties along the line-of-sight, e.g. the MBB dust temperature is the not the mass-weighted temperature of all the dust particles contributing to the measured emission. Rather, the MBB parameters are simply the empirical parameters which best fit the overall dust emission. This approach is well-suited for component separation, however for an improved understanding of the thermal dust emission mechanism itself more complex models are required (e.g. Draine & Li (2007); Compiègne et al. (2011); Jones et al. (2013)).

Planck Collaboration et al. (2014) introduce a two-stage approach to the customary pixel-by-pixel fit. First they smooth the $N_{\text{side}} = 2048$ HFI maps to 30 arcmin and fit a MBB with three free parameters to obtain the spectral indices per pixel. Next they fit the 5 arcmin (essentially full resolution) HFI maps, fixing the spectral index values to those obtained at 30 arcmin and allowing two free parameters for the fit. The result being all-sky maps of dust temperature and 353 GHz optical depth at 5 arcmin and spectral index at 30 arcmin. Smoothing the data averages out the CIB anisotropies to a greater extent than it averages out the spatial correlations of thermal

★ E-mail: melis.irfan@cea.fr

dust emission; the CIB has a flatter angular power spectrum than thermal dust. However the effects of smoothing on thermal dust emission are not negligible; Planck Collaboration et al. (2014) note the trade-off between sensitivity to small-scale variations of thermal dust emission and smoothing for component separation purposes. The all-sky maps of thermal dust emission formed from the three fitted parameters display variations of 30 per cent, at all scales, from the Finkbeiner, Davis and Schlegel thermal dust model (Finkbeiner et al. 1999).

Planck Collaboration et al. (2016c) presents an improvement on the Planck Collaboration et al. (2014) method by only smoothing the data within the regions where the CIB anisotropies start to dominate the total signal at small angular scales. Their analysis is performed within the needlet (spherical wavelet) domain, where emissions are modelled as transient waveforms within pixel and harmonic space. Planck Collaboration et al. (2016c) account for the ‘nuisance’ (non-thermal dust) contributions using a covariance matrix formed from noise, CIB and CMB estimates. This nuisance covariance matrix is used to identify regions of the sky where thermal dust emission dominates over CIB, CMB and noise. Within the nuisance dominated regions the data are smoothed and reassessed, if the nuisance term continues to dominate the width of the Gaussian kernel used for smoothing is increased by a factor of two and the process repeated. Therefore the Planck Collaboration et al. (2016c) all-sky estimates of thermal dust emission have a range of effective beam sizes from $\text{FWHM} = 5.0$ arcmin, within the high signal-to-noise regions (over 65 per cent of the sky) to $\text{FWHM} = 21.8$ arcmin at high Galactic latitudes. The Planck Collaboration et al. (2016c) thermal dust maps are shown to be an improvement on the Planck Collaboration et al. (2014) maps when the residual CIB maps from both methods and the reduced χ^2 from the MBB fits are compared.

1.1 Limitations and objectives

Both the Planck Collaboration et al. (2014) and Planck Collaboration et al. (2016c) method have the disadvantage of being computationally intensive, as they perform the MBB fit on each pixel for the full resolution *Planck* data (over fifty million pixels). While the Planck Collaboration et al. (2014) method has the ability to provide thermal dust temperature and 353 GHz optical depth maps at the full 5 arcmin resolution, any variations in the dust spectral index at angular scales smaller than 30 arcmin are smoothed over. Contrarily though, it could be argued that the Planck Collaboration et al. (2014) maps are not smoothed enough as strong evidence of CIB contamination still remain. Therefore it is clear that smoothing all pixels within the full sky HFI maps to the same extent is not optimum. The strength of the Planck Collaboration et al. (2016c) method is that it locates the areas worst affected by CIB contamination and applies the largest degree of smoothing within these areas. As successful as this method is it still presents an incomplete picture of the spatial variance of the dust parameters across the sky, due to smoothing. We aim to provide a method of parameter estimation for thermal dust emission, contaminated by CIB and noise contributions, which can be applied to all-sky data at full *Planck* resolution, is robust to a wide range of signal-to-noise ratios across the sky and is faster than a pixel-by-pixel MBB fit.

We present a new dust parameter estimation method, which we refer to as **premise**: Parameter Recovery Exploiting Model Informed Sparse Estimates. This novel approach builds upon a two-stage procedure:

- a filtering technique that first performs CIB and noise removal.

In the spirit of GNILC, it makes use of the nuisance covariance matrix but further exploits the natural sparsity of the dust emission in the wavelet domain.

- an innovative parameter estimation algorithm that builds upon recent advances in applied mathematics to find the optimum parameter values per pixel. This achieved using the least squares estimator on the residual between the model maps and the empirical data (thermal dust plus nuisance terms) within the wavelet domain and adding a penalisation factor to favour sparsity.

As **premise** never runs a pixel-by-pixel MBB fit it is faster than the traditional methods and the use of sparsity in the second stage reduces the need for smoothing whilst still providing an improvement in parameter estimation within noise dominated regions.

Starting with simulation data representing the combination of thermal dust emission, point sources, instrumental noise, CMB and CIB we apply **premise** to produce estimates of the thermal dust MBB parameters. The simulated thermal dust emission was created using a single MBB with three parameters: temperature, spectral index and optical depth at 353 GHz. Fitting the same model to the **premise** thermal dust emission maps provides a comparison between the fitted and the true parameters which can be used to evaluate **premise**. Additionally, we implement a GNILC methodology (Planck Collaboration et al. 2016c) to enable comparisons between GNILC and **premise**.

1.2 Notations

Throughout this work we make use of numerous notations, we summarise them here for clarity. $x_{\nu_i}[k]$ indicates the total flux for pixel k at frequency ν_i and is the combination of CMB, CIB and instrumental noise, point sources and thermal dust emission:

$$x_{\nu_i}[k] = x_{\nu_i}[k]^{\text{CMB}} + x_{\nu_i}[k]^{\text{CIB}} + x_{\nu_i}[k]^{\text{noise}} + x_{\nu_i}[k]^{\text{ps}} + x_{\nu_i}[k]^{\text{dust}}. \quad (1)$$

When considering all of the observational frequencies simultaneously the vector form of the above equation can be used:

$$\mathbf{X}[k] = X[k]^{\text{CMB}} + X[k]^{\text{CIB}} + X[k]^{\text{noise}} + X[k]^{\text{ps}} + X[k]^{\text{dust}} \quad (2)$$

and for all frequencies and all pixels we have the matrix form:

$$\mathbf{X} = \mathbf{X}^{\text{CMB}} + \mathbf{X}^{\text{CIB}} + \mathbf{X}^{\text{noise}} + \mathbf{X}^{\text{ps}} + \mathbf{X}^{\text{dust}}. \quad (3)$$

The forward wavelet transformation of single frequency map is denoted as $x_{\nu_i}\Phi$, with $0 < j < J$ being the number of the wavelet scale. At each scale the transformations produces 2^{j-1} coefficients which consist of the coarse scale (c) and wavelet coefficients (w):

$$x_{\nu_i}\Phi = \alpha_{\nu_i} = [c_{\nu_i}, w_{\nu_i}^{(j=1)} \dots w_{\nu_i}^{(j=J)}]. \quad (4)$$

The L1-norm of a variable in the wavelet domain is equivalent to the sum of the wavelet coefficients:

$$\|x_{\nu_i}\Phi\|_{\ell_1} = \sum_j |w_{\nu_i}^j| \quad (5)$$

Formally, the wavelet decomposition of the data matrix \mathbf{X} yields the following decomposition at each scale j :

$$\mathbf{W}_j = \mathbf{W}_j^{\text{CMB}} + \mathbf{W}_j^{\text{CIB}} + \mathbf{W}_j^{\text{noise}} + \mathbf{W}_j^{\text{ps}} + \mathbf{W}_j^{\text{dust}} \quad (6)$$

where the rows of the matrix \mathbf{W}_j contain the j -th wavelet scale of each of the channels.

S_τ stands for the soft-thresholding operator with threshold τ , which is described as follows for some value u :

$$S_\tau(u) = \begin{cases} u - \tau \text{sign}(u) & \text{if } |u| > \tau \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

When referring to an iterative process we denote a variable at a particular time ‘t’ as $x^{(t)}$ and specifically for gradient descent iterations we define the minimum of said variable as $x^{\frac{1}{2}}$.

The paper is organised as follows: section 2 introduces the simulation data alongside the ancillary data used, section 3 details the steps within `premise` and section 4 presents our results in comparison with results obtained via the GNILC methodology.

2 DATA AND PREPROCESSING

2.1 Simulated Thermal Dust

The data used in this paper are simulation data, formed from equation (8). The GNILC¹ all-sky temperature and spectral index maps provide the required temperature (T) and spectral index (β), the *Planck* FFP8 (Planck Collaboration et al. 2016b) thermal dust map at 353 GHz provides the normalisation factor and B denotes the blackbody function.

$$x_{\nu_i}^{\text{dust}} = x_{353 \text{ GHz}}^{\text{ffp8}} \times \frac{B(T, \nu)}{B(T, 353 \text{ GHz})} \times \left(\frac{\nu}{353 \text{ GHz}} \right)^\beta \times \text{cc}^{-1}. \quad (8)$$

The single spectral index model was preferred over the two-index model (Meisner & Finkbeiner 2015) as the two-index model has been shown to improve the fit to the data when frequencies below 353 GHz are used. Equation (8) was evaluated at frequencies 353, 545, 857 and 3000 GHz to emulate *Planck* HFI data used in combination with the 100 μm IRIS map (Miville-Deschénes & Lagache 2005). These fluxes were multiplied by their inverse colour corrections (cc^{-1}) to further align them with real data. The colour corrections were calculated across the *Planck* HFI bandpasses for frequencies 353, 545 and 857 GHz. The 3000 GHz colour correction factors were taken from Table 3 of Miville-Deschénes & Lagache (2005).

Simulation data were not created for the lowest two HFI frequencies (100 and 217 GHz) following the method of (Planck Collaboration et al. 2016c), which fits a MBB model to the GNILC 353, 545, 857 and 3000 GHz thermal dust estimates. The 353 GHz FFP8 data were neither smoothed nor downgraded, so all our simulation data are at N_{side} 2048 with a FWHM of ~ 5 arcmin.

2.2 Simulated CIB, CMB, noise and point sources

The simulated total flux maps used differ for the GNILC methodology and `premise`. While GNILC includes the CMB within its covariance matrix of nuisance terms, `premise` is intended for use on CMB subtracted maps. In Bobin et al. (2016), we introduced a sparsity-based component separation method coined L-GMCA to estimate a precise CMB map from the *Planck* data. More specifically, we showed that the estimated CMB map has very low foreground contamination. Consequently, the L-GMCA CMB map will be removed prior to applying the `premise` algorithm.

The total flux maps used for the GNILC methodology were constructed as:

$$\mathbf{X}^{\text{GNILC}} = \mathbf{X}^{\text{dust}} + \mathbf{X}^{\text{CIB}} + \mathbf{X}^{\text{CMB}} + \mathbf{X}^{\text{noise}} + \mathbf{X}^{\text{ps}}, \quad (9)$$

where the FFP8 simulations provide the instrumental noise, point sources and CMB contributions. A 3000 GHz CIB map of N_{side} 2048 and FWHM 5 arcmin was created using the methodology detailed in Appendix C of Planck Collaboration et al. (2014). Gaussian noise with a median level of 0.06 MJy sr $^{-1}$ (Planck Collaboration et al. 2014) was used for the 3000 GHz noise map. No CMB nor point source contributions were included at 3000 GHz. The total flux maps used for `premise` were constructed as:

$$\mathbf{X}^{\text{premise}} = \mathbf{X}^{\text{dust}} + \widetilde{\mathbf{X}^{\text{CIB}}} + \mathbf{X}^{\tilde{n}}(\nu) + \mathbf{X}^{\text{ps}}, \quad (10)$$

where $\widetilde{\mathbf{X}^{\text{CIB}}}$ and $\mathbf{X}^{\tilde{n}}(\nu)$ are simulated CIB and instrumental noise maps featuring fractional CMB emission, representative of the CIB and noise properties of an intensity map after the L-GMCA CMB subtraction.

2.3 LAB data and the large-scale CIB contribution

The CIB is seen to contribute to the overall measured intensity in two ways: small-scale variations and a large-scale intensity which manifest as a constant, additive offset to the pure thermal dust intensity. Fig. 1 shows a 1D slice of a patch of data at 353 GHz. The green line shows the total emission while the blue line shows the pure thermal dust emission. The large-scale CIB contribution can clearly be seen as a positive offset while the small-scale CIB contributions, alongside instrumental noise, are seen as Gaussian-like variations around the thermal dust mean level.

Before the simulated maps of total emission can be processed the large-scale CIB offsets must first be subtracted. The removal of the large-scale CIB offset is achieved through the use of LAB H α Survey data (Kalberla et al. 2005). These data are available in the form of all-sky HEALPix maps (Górski et al. 2005) at N_{side} 512 and a FWHM of ~ 36 arcmin.² The constant CIB offsets for each frequency were calculated using the method described in Planck Collaboration et al. (2014), see Appendix A for details. The resulting offsets were found to be 0.126, 0.331, 0.641 and 0.657 MJy sr $^{-1}$ for 353, 545, 857 and 3000 GHz respectively. These values were subtracted from the total flux maps. The mean values for the simulated CIB maps are 0.140, 0.366, 0.718 and 0.718 MJy sr $^{-1}$, placing an 8.5 – 11 per cent error on the linear regression method.

2.4 GNILC and `premise` preprocessing

We implement two algorithms: our own, `premise` and a GNILC-like methodology. The principal steps for both algorithms are summarised in Fig. 2: removing large-scale CIB offsets, filtering out noise and small-scale CIB anisotropies and fitting a MBB. GNILC makes use of a covariance matrix of the combined nuisance terms to locate dimensions within signal subspace where the desired signal, thermal dust emission in this case, is dominant. The nuisance terms are the small-scale CIB contributions, the CMB and instrumental noise. As this paper is concerned with simulation data the exact CIB, CMB and instrumental noise maps used to make the total flux maps are given as nuisance term estimates, therefore the estimation of the nuisance covariance matrix is unrealistically perfect. The GNILC

¹ <http://pla.esac.esa.int/pla/>

² <https://lambda.gsfc.nasa.gov>

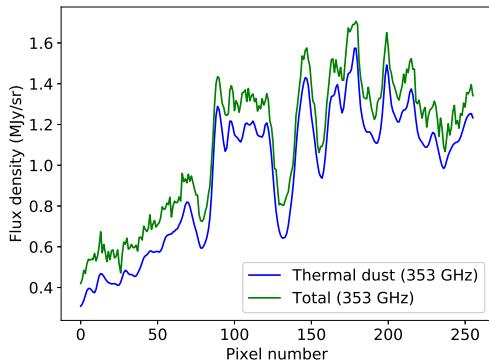


Figure 1. One dimensional slice of a patch of data at 353 GHz. The green line represents thermal dust plus CIB plus instrumental noise, the blue line is just thermal dust.

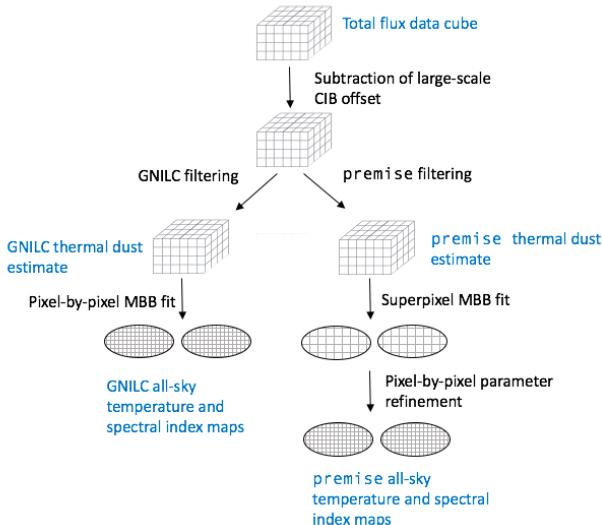


Figure 2. The principal steps of the both the GNILC and premise algorithms.

methodology for filtering total flux maps to produce pure thermal dust estimates that we implement is not identical to that described in [Planck Collaboration et al. \(2016c\)](#), therefore we detail our GNILC filtering technique in Appendix B, highlighting any differences.

After the initial filtering the following MBB model is fit pixel-by-pixel to the GNILC estimate of pure thermal dust emission:

$$x_{\nu_i} = \tau_{353} \times B(T, \nu) \times \left(\frac{\nu}{353 \text{ GHz}} \right)^{\beta} \times \text{cc}^{-1}, \quad (11)$$

where the optical depth at 353 GHz (τ_{353}), the dust temperature (T) and the thermal dust spectral index (β) are the free parameters in the fit. The inverse colour correction (cc^{-1}) also depends on T and β .

From Fig. 2 it can be seen that both GNILC and premise filter the total flux maps to produce thermal dust estimates. For GNILC, the principal aim is to construct all-sky thermal dust estimates which can then provide values for the MBB parameters through a pixel-by-pixel fit. The goal of premise is the inverse; the premise filtered dust estimates are only of use as inputs to the MBB fit, our aim is to calculate the thermal dust MBB parameters as accurately as possible.

2.5 Dealing with point sources

Point sources were removed from the 353, 545 and 857 GHz data using the appropriate FFP8 masks for each frequency. The premise algorithm is capable of processing masked data as the use of the discrete wavelet transform allows for the reconstruction of the full wavelet, providing the Shannon sampling condition is fulfilled. The GNILC methodology, however, requires each data pixel to provide astrophysical information. Therefore the masked GNILC data were inpainted using a morphological component analysis technique ([Abrial et al. 2007](#)) applied to each frequency map independently.

3 THE premise METHOD

Estimating the thermal dust parameters per pixel from the *Planck* data raises two important issues: i) it requires solving large-scale (*i.e.* over fifty million pixels for full-resolution *Planck* data) non-linear parameter estimation problems and ii) it is highly sensitive to nuisance terms such as noise, CIB and CMB, which explains why standard pixel-based fitting performs poorly in noise dominated regions of the sky. The method of filtering nuisance contributions prior to applying pixel-based fitting has two limitations: i) it only uses statistical information about the nuisance term but does not account for the statistics of the dust to be retrieved, and ii) pixel-based fitting will still be sensitive to dust estimation biases induced by the dust filtering procedure. For that purpose, the proposed premise method is a new parameter estimation procedure that minimises a non-linear least-square cost function to which an extra penalisation is added to favour sparse parameter maps in the wavelet domain. This procedure allows for a fast and effective full-sky estimation and accounts for the naturally sparse distribution of the temperature and spectral index maps in the wavelet domain.

[Planck Collaboration et al. \(2014\)](#) split their MBB fit into two stages to combat parameter degeneracies introduced by the CIB anisotropies, choosing first to calculate the thermal dust spectral indices and then the temperature and optical depth at 353 GHz. We too experienced such degeneracies and so also opt for a two stage approach, choosing to first calculate T and β and then, through the introduction of a dust template, we attempt to break the parameter degeneracies and calculate τ_{353} .

3.1 Fitting T and β from the MBB model

Fitting for the spectral index and the temperature in the MBB model requires dealing with the non-linear relationship between these parameters and the observed pixel. For a given pixel, this reads as:

$$\min_{\beta, T} \sum_i \left(x_{\nu_i}[k] - x_{\nu_i}[k]^{\text{dust}}(\beta[k], T[k]) \right)^2 \quad (12)$$

In premise and unlike currently available pixel-based fitting methods, we propose estimating full-sky maps of the parameters while accounting for their strong correlation across pixels. This further requires solving a minimisation problem of the form:

$$\min_{\beta, T} \lambda_b \|\beta\Phi\|_{\ell_1} + \lambda_t \|T\Phi\|_{\ell_1} + \sum_{i,k} \left(x_{\nu_i}[k] - M_{\nu_i}[k] x_{\nu_i}[k]^{\text{dust}}(\beta[k], T[k]) \right)^2 \quad (13)$$

where the quadratic term measures the discrepancy between the observations and the MBB model. Point source masks ($M_{\nu_i}[k]$) take the value of 1 when no source is present and 0 otherwise. The penalisation terms enforce the sparsity of the estimated T and β in

the wavelet domain through the sum of the wavelet coefficients. In the next section, we will make use of the undecimated and isotropic wavelet transform (Starck et al. 2007).

We will show in section 3.1.2 how such a penalised non-linear least-square minimisation problem can be solved using an iterative thresholding algorithm. However, since the MBB model is non-linear, these algorithms can be sensitive to the initial point from which the algorithm starts. To alleviate this issue, we propose a two-stage fitting approach that is composed of: i) a fast initialisation procedure that builds upon a quadtree decomposition of the data in the wavelet domain, and ii) a refinement stage based on an iterative thresholding algorithm.

3.1.1 Fast wavelet-based initialisation

The goal of this initialisation step is to provide a first estimate of the dust emission that is robust to CIB and noise, with low computational cost. For that purpose, and similarly to GNILC, a model fitting is performed on CIB and noise filtered data. In contrast to pixel-by-pixel fitting, the proposed initialisation procedure makes use of the quadtree technique, which significantly reduces the computational cost of this procedure. This eventually yields a trade-off between good estimation accuracy and robustness to noise and CIB but too a lesser extent than smoothing.

Disentangling between the dust emission and nuisance contamination The filtering step aims at improving the robustness of the initialisation step with respect to noise and CIB. We would like to highlight that the filtered data are only used during the initialisation; the final estimate of the dust estimation is performed on the raw data. Filtering consists of disentangling between the dust emission and the nuisance components, this can be achieved thanks to i) the Gaussianity of the CIB and noise components and ii) the sparsity of the dust emission in the wavelet domain.

The GNILC filtering requires an estimation of the nuisance terms from which to calculate the nuisance covariance matrix from. For our implementation of GNILC the CIB, instrumental noise and CMB estimates are in fact the actual CIB, instrumental noise and CMB contributions to the total flux. As *premise* is intended for use after subtraction of the CMB we too use the same CIB and instrumental noise estimates as our GNILC implementation but this differs to the actual CIB and instrumental noise present in our total flux maps (see section 2.2).

One of the major advantages of GNILC is the ability to exploit the statistical properties of CIB and noise to disentangle between these components and the dust emission. Hence, the filtering stage will be based on a similar data processing with three major differences:

- As it is described in Appendix B, GNILC identifies the dust emission subspace by thresholding the eigenvalues of the data covariance matrix in the needlet domain after nuisance-whitened. For that purpose, it relies on a dimensionality selection criterion called AIC (Akaike Information Criterion), which turns out to be slightly conservative in practice. Consequently, the AIC tends to provide an over-smoothed estimation of the filtered data. In the proposed approach, the threshold is based on the actual statistics of the eigenvalues of the nuisance covariance. Since both the CIB and noise follow Gaussian distributions with known covariance matrices, their eigenvalues after whitening follow a Marcenko-Pastur distribution (Mehta 2004). The dust emission subspace is estimated

by identifying the eigenvalues which exceed a particular threshold, so chosen to ensure that the false detection probability is lower than 10^{-3} according the Marcenko-Pastur distribution. This procedure is similar in spirit to the 3σ thresholding used in standard Gaussian statistics.

- GNILC further relies on a Gaussian smoothing of the data covariance matrices, which dramatically increases the computational cost of the filtering step. Our filtering is performed on overlapping patches or super-pixels with an overlapping ratio of 0.5. This results in a decrease in the computational time required for this step without significantly altering the filtering quality.
- The proposed GNILC-based filtering only exploits the statistical properties of CIB and noise; it does not profit from the sparsity of dust emission in the wavelet domain. Additionally, since it relies on a spatial smoothing of the data, it yields a smooth estimate ($\tilde{\mathbf{X}}^{\text{filt}}$) of the dust emission with lower spatial resolution. We attempt to mitigate the effects of smoothing by exploiting the sparsity of the dust emission in the wavelet domain. Since neither CIB nor noise have a sparse representation in the wavelet domain, we propose to enhance the filtering stage by updating the above filtered dust with sparse deviations in the wavelet domain. These sparse deviations can only originate from dust emission. Formally, each wavelet band of the estimated dust emission will be obtained as $\tilde{\mathbf{W}}_j^{\text{dust}} = \tilde{\mathbf{W}}_j^{\text{filt}} + \Delta_j$. The sparse deviation term Δ_j is computed at each wavelet scale by solving a minimisation that is analogous to a sparse denoising problem (Starck et al. 2010):

$$\min_{\Delta_j} \lambda_j \|\Delta_j\|_{2,1} + \frac{1}{2} \|\mathbf{W}_j - \tilde{\mathbf{W}}_j^{\text{filt}} - \Delta_j\|_F^2, \quad (14)$$

where \mathbf{W}_j is the raw data and $\|\cdot\|_F^2$ is the Frobenius norm. The L2,1-norm is defined for some matrix \mathbf{Y} as:

$$\|\mathbf{Y}\|_{2,1} = \sum_k \sqrt{\sum_i Y_{vi}[k]^2}. \quad (15)$$

Unlike the L1-norm, this regularisation term allows us to enforce the column-wise sparsity of the data, which is adapted to capture signals that have sparse distributions in space and are likely to be present in all channels. The problem admits a solution with a closed-form expression (Kowalski 2009) so that at each pixel:

$$\Delta_j[k] = \begin{cases} (\mathbf{W}_j[k] - \tilde{\mathbf{W}}_j^{\text{filt}}[k]) \times \left(1 - \frac{\lambda_j}{\|\mathbf{W}_j[k] - \tilde{\mathbf{W}}_j^{\text{filt}}[k]\|_2^2}\right) & \text{if } \lambda_j < \|\mathbf{W}_j[k] - \tilde{\mathbf{W}}_j^{\text{filt}}[k]\|_2^2 \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

The solution is therefore nonzero whenever the χ^2 of the residual $\mathbf{W}_j[k] - \tilde{\mathbf{W}}_j^{\text{filt}}[k]$ is lower than the value λ_j and equal to a pruned version of the residual otherwise. Since λ_j plays the role of a threshold on the residual it can be chosen based on the statistical significance of the measured χ^2 . At each scale λ_j is fixed to 2.7, which corresponds to a p-value of 10 per cent.

Combining both a GNILC-based filtering and wavelet-based filtering enables us to account for the statistical properties of the CIB and noise components as well as the sparsity of the dust emission in the wavelet domain.

Quadtree-based fitting in the wavelet domain Our fast, initial parameter estimation technique makes use of a quadtree. A quadtree recursively divides the given data into quarters until a particular

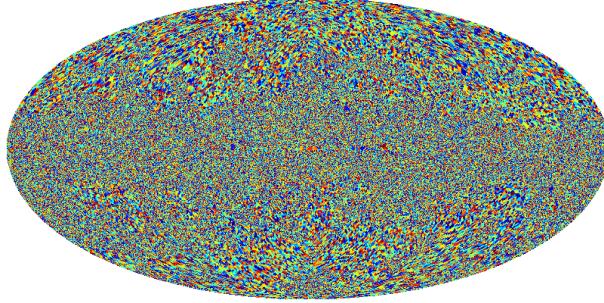


Figure 3. All-sky map of the different patches chosen by the quadtree.

criterion is no longer achieved within the data patch. The quadtree used in this method recursively divides a square of data until the number of data points (in this case, reduced χ^2 values) with a value greater than 2 within the patch are less than 10 per cent of the total number of data points. Fig. 3 shows the patches selected by the quadtree over the full sky. The patches sizes are visibly smaller within and near to the Galactic plane where the total flux across neighbouring pixels is less consistent than at high latitudes. Very close to the Galactic centre several patch sizes larger than their neighbouring patches can be seen: these areas contain so many masked pixels (due to point sources) that the quadtree is forbidden from dividing the region into smaller areas.

The premise fitting process works on each of the twelve 2048X2048, HEALPix 2D faces which make up the full sky as follows:

1. The noise covariance matrix (\mathbf{R}_n) of the full data face is calculated within the wavelet domain:

$$\mathbf{R}_n^j = \frac{1}{2048^2} (\mathbf{W}_j^n (\mathbf{W}_j^n)^T), \quad (17)$$

where

$$\mathbf{W}_j^n = \mathbf{W}_j^{\text{noise}} + \mathbf{W}_j^{\text{CIB}} - \mathbf{W}_j^{\text{CIB offset}}$$

2. The region is split into patches of 128X128
3. Each patch is treated as a super-pixel: the mean flux density for each frequency represents the whole patch at that frequency
4. A MBB fit to each super-pixel yields the parameters required to form a thermal dust estimate for each frequency at that super-pixel.
5. The data-model residual is calculated per original pixel:

$$\text{Residual}[k] = \text{Total Flux}[k] - \text{Estimated thermal dust flux}[k]$$

6. This residual matrix (which has dimensions of number of frequencies by number of pixels) is transformed into the wavelet domain (\mathbf{W}_j^r).

7. The reduced χ^2 at each of the wavelet scales is calculated for each pixel:

$$\chi_{\text{red}}^2 = (\mathbf{W}_j^r)^T (\mathbf{W}_j^n)^{-1} \mathbf{W}_j^r \quad (18)$$

As there is only one degree of freedom the reduced χ^2 is just the χ^2 .

8. The reduced χ^2 is fed back into the quadtree to split the data into patches where T and β are constant enough to be well characterised by a MBB with a single value of T and β across all pixels within the patch. The final patches chosen are the quadtree determined super-pixels.

9. The fit is re-run fit on the quadtree chosen super-pixels and initial maps of T and β estimates are obtained.

Tuning the quadtree parameters

- Wavelet scales: For steps 1. and 5. and 7. the number of wavelet scales used is four. For step 8. the higher wavelet scales are used for larger patch sizes so when deciding whether or not to split a 64X64 patch the quadtree uses the reduced χ^2 values from the fourth wavelet scale, for a 16X16 patch the first wavelet scale is used.
- Patch area: A lower limit of 8 is enforced to ensure that the quadtree cannot split the data into patches smaller than 8X8 pixels. The quadtree is also prohibited from splitting a patch into quarters if one or more of those quarters contains only masked data.

Running the MBB fit on super-pixels, as opposed to the true pixels, reduces the computational time by, at minimum, a factor of 64 (the minimum patch area being 64 pixels).

3.1.2 Parameter refinement

In this section, we detail how the MBB parameters are estimated from the raw data by solving the penalised least-square estimation problem in Equation 13. The main difficulty lies in the non-differentiability of the L1-norm. Consequently, minimising the cost function (Eq. 19) cannot be achieved using standard gradient descent methods. The recently introduced proximal algorithms (Parikh et al. 2014) can provide the perfect framework to design an effective minimiser for the problem. We will make use of the FISTA algorithm: an iterative projected gradient descent algorithm introduced in Beck & Teboulle (2009) which has been specifically designed to solve penalised least-square problems with linear models. Projected gradient descent techniques have also shown to be well-suited for tackling, not only linear but also non-linear problems (Teschke & Bories 2010), making them a perfect fit to solve Eq. 19.

The parameter optimisation step of premise builds upon an iterative estimation procedure, composing of the following two stages:

- *Gradient descent*: this step consists in performing a single gradient descent step of the data fidelity term with respect to β and T :

$$\begin{aligned} \beta^{(\frac{1}{2})}[k] &= \beta^{(t)}[k] + \\ &\alpha g_{\beta}^T[k] \mathcal{M}[k] (X[k] - \mathcal{M}[k] X[k]^{\text{dust}}(\beta^{(t)}[k], T^{(t)}[k])), \\ T^{(\frac{1}{2})}[k] &= T^{(t)}[k] + \\ &\alpha g_T^T[k] \mathcal{M}[k] (X[k] - \mathcal{M}[k] X[k]^{\text{dust}}(\beta^{(t)}[k], T^{(t)}[k])), \end{aligned} \quad (19)$$

where α is the gradient path length and $g_Y[k]$ is a vector whose i -th is equal to the derivative of the MBB model:

$$\frac{\partial X_{Y_i}^{\text{dust}}[k]}{\partial Y[k]}(\beta[k]^{(t)}, T[k]^{(t)}),$$

where Y indicates what the derivative is in respect to (in this case β or T). These quantities are computed analytically from the expression of the MBB model.

- **Projection step:** In this step, the estimated β and T maps are thresholded in the wavelet domain:

$$\beta^{(t+1)}[k] = S_{\alpha \lambda_B} \left(\beta^{(\frac{1}{2})} \Phi \right) \Phi^{-1}, \quad (20)$$

$$T^{(t+1)}[k] = S_{\alpha \lambda_T} \left(T^{(\frac{1}{2})} \Phi \right) \Phi^{-1}. \quad (21)$$

Initialise $\beta^{(0)}$ and $T^{(0)}$ with the quad-tree based fitting technique,

At each iteration t .

1 - Gradient descent step for each pixel

$$\begin{aligned} \beta^{(\frac{1}{2})}[k] &= \beta^{(t)}[k] + \\ &\alpha g_\beta^T[k] \mathcal{M}[k] (X[k] - \mathcal{M}_{\nu_i}[k] X^{\text{dust}}[k](\beta^{(t)}[k], T^{(t)}[k])) \\ T^{(\frac{1}{2})}[k] &= T^{(t)}[k] + \\ &\alpha g_T^T[k] \mathcal{M}[k] (X[k] - \mathcal{M}_{\nu_i}[k] X^{\text{dust}}[k](\beta^{(t)}[k], T^{(t)}[k])) \end{aligned}$$

2 - Thresholding step

$$\begin{aligned} \beta^{(t+1)}[k] &= S_{\alpha \lambda_B} \left(\beta^{(\frac{1}{2})} \Phi \right) \Phi^{-1} \\ T^{(t+1)}[k] &= S_{\alpha \lambda_T} \left(T^{(\frac{1}{2})} \Phi \right) \Phi^{-1} \end{aligned}$$

Stop when a given criterion is valid.

Tuning the refinement parameters

- Gradient path length: The algorithm converges to the stationary point of the problem in Eq. 19 provided that $\alpha \leq \min_k 1/\|H[k]\|_2$, where H is the Hessian matrix of the data fidelity term with respect to β and T . The spectral norm, $\|H[k]\|_2$, is the largest singular value of the Hessian matrix. The proposed iterative thresholded gradient descent technique builds upon a local linear approximation of the data fidelity term about the current estimate at iteration t . The same linear approximation yields $\alpha \leq \min_k 1/(g_\beta^T[k] g_\beta[k] + g_T^T[k] g_T[k])$, which is computed analytically at each iteration.
- Thresholds: A careful choice of the thresholds λ_B and λ_T are essential since these thresholds tune the trade off between the data fidelity term and the sparse penalisation. For that purpose, we make use of a very effective, heuristic approach to choose these parameters automatically. The goal of the thresholding stage is to save the wavelet coefficients with significant amplitude while removing noise or non-sparse elements. In this context, the standard deviation of the non-sparse elements can be computed using the MAD (Median Absolute Deviation) (Starck et al. 2010); the thresholds are then chosen as $\lambda_B = 2\sigma_{\text{MAD}(\beta^{(\frac{1}{2})})}$ and $\lambda_T = 2\sigma_{\text{MAD}(T^{(\frac{1}{2})})}$. In practice, the thresholds are computed independently at each of the five wavelet scales.
- Stopping criterion: The proposed iterative algorithm stops when the relative variation of the estimates between two consecutive

iterations is lower than a given level:

$$\max_{\beta, T} \left(\frac{\|\beta^{(t+1)}[k] - \beta^{(t)}[k]\|_{\ell_2}}{\|\beta^{(t)}[k]\|_{\ell_2}}, \frac{\|T^{(t+1)}[k] - T^{(t)}[k]\|_{\ell_2}}{\|T^{(t)}[k]\|_{\ell_2}} \right) \leq 10^{-6}$$

3.1.3 Optical depth at 353 GHz

Up to this point we have only dealt with two of the three MBB parameters: temperature and spectral index. Both the fast wavelet-based initialisation and the parameter refinement steps can be utilised to produce optical depth values just as they do for the temperature and spectral index. However, due to the degeneracies between the three parameters introduced through residual CIB and noise, `premise` produces less accurate optical depth estimates than desired. Both Planck Collaboration et al. (2014) and Planck Collaboration et al. (2016c) encounter these degeneracies and deal with them through smoothing; in Planck Collaboration et al. (2014) the spectral index maps are smoothed to 30 arcmin while in Planck Collaboration et al. (2016c) all three parameters are smoothed within the lowest signal-to-noise regions of the sky.

`premise` is specialised to trace the spatial variations of the dust temperature and spectral index, to the detriment of the overall normalisation factor (the optical depth). At this point it is therefore useful to introduce additional information to break the parameter degeneracies and recover the normalisation factor: a thermal dust template. The total flux 857 GHz map is known to have minimal CIB contamination and so provides us with the opportunity to recover the 353 GHz optical depth at 5 arcmin using the `premise` estimates for temperature and spectral index:

$$\tau_{353} = \frac{x_{857} \times cc}{B(T, 857 \text{ GHz}) \times \left(\frac{\nu}{353 \text{ GHz}} \right)^\beta}. \quad (22)$$

As the total flux data include point sources, a masked and inpainted (Abrial et al. 2007) version of the total flux map is used to recover the optical depth at 353 GHz.

4 RESULTS

4.1 Regions

We compute `premise` on the full sky, however as GNILC is a computationally intensive algorithm we only run the GNILC methodology on four test regions and conduct our comparison within these four, 256X256 pixel regions. The regions comprise of a high signal-to-noise region in the Galactic plane (region 1), two medium signal-to-noise regions at intermediate latitudes - one at a central longitude (region 2) and one at zero longitude (region 3) and a low signal-to-noise polar region (region 4). The location of these regions within a HEALPix sphere is shown in Fig. 4.

4.2 Comparison of parameter estimation within Regions 1–4

In the following section we compare the GNILC and `premise` fitted τ_{353} , T and β values for Regions 1–4 to the true τ_{353} , T and β . Fig. 5 shows histograms of the actual differences between the true parameter values and those derived from the GNILC methodology/`premise`. Signal-to-noise maps (the ratio of pure thermal dust emission to the combination of instrumental noise and CIB) of each region are shown in the right-hand column. For region 1, the region with the largest signal-to-noise ratio, the GNILC τ_{353} and T parameters show smaller differences to the true parameter values than

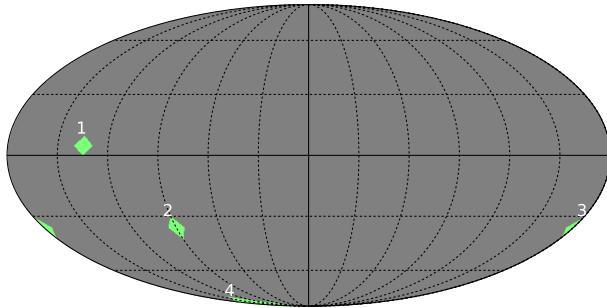


Figure 4. The location on the sphere of the four 256X256 pixel regions chosen for analysis.

premise. The opposite being true for the spectral index estimates. GNILC has an advantage over **premise**, at high signal-to-noise, with regards to the estimation of tau as the **premise** quadtree specialises in detecting spectral deviations between the MMB model and the data as opposed to selecting super-pixels sizes to best estimate the normalisation factor.

Outside of the Galactic plane, however, smaller differences between the true and fitted parameters can be seen for **premise**. This can be seen most clearly in regions 2 and 4, were the thermal dust signal-to-noise ratio is lowest. As the signal-to-noise drops the GNILC algorithm increases the degree of smoothing and their parameter estimates deviate further from the true values. Therefore, from the perspective of parameter estimation across a diverse range of signal-to-noise ratios, **premise** provides the more robust set of parameter values for the full sky.

Fig. 6 shows maps of the absolute percentage differences between the true parameter values and those derived from the GNILC methodology/**premise**. For region 1, the pixels with the poorest parameter estimates (for both methods) are those surrounded by clusters of extragalactic point sources. Fig. 7 is Fig. 6 a) with the total point source mask (the individual frequency mask maps multiplied together) applied. GNILC uses inpainting to determine β and T within masked regions while **premise** just averages over larger super-pixel sizes than desirable for those regions where large portions of the data are masked out. As a result of this, the **premise** parameter estimates suffer most within regions of high point-source contamination, like the Galactic plane.

Within regions 2, 3 and 4, Fig. 6 reenforces the fact that the GNILC estimations suffer from over-smoothing as whole patches of the parameter space are misestimated. Region 4 is plotted using the same colour-scale as the other regions to highlight how poorly both methods fair (when compared to their optimal performances) within the lowest signal-to-noise regions of the full sky.

4.3 Full sky **premise**

4.3.1 Parameters

The all-sky **premise** thermal dust temperature, spectral index and optical depth at 353 GHz are shown in Fig. 8. Fig. 9 displays the absolute differences between true and **premise** calculated parameters. From the difference maps it is clear that the estimation of both β and T suffer within the Galactic plane and at high latitudes. The result at high latitudes is unsurprising as the thermal dust signal-to-

Table 1. The percentage difference between true and **premise** estimated thermal dust parameters for various percentiles.

Value	Median % Δ	1 σ % Δ	2 σ % Δ	3 σ % Δ
Temperature	1.7	2.8	8.0	16.5
Spectral index	3.4	5.7	15.4	25.6
Optical depth (353 GHz)	3.7	7.2	31.2	77.0
353 GHz emission	7.9	9.4	29.4	77.4
545 GHz emission	7.3	8.5	23.9	65.2
857 GHz emission	2.2	4.6	24.2	64.8
3000 GHz emission	7.4	10.9	33.7	89.4

noise ratio is at its lowest within these regions. Yet **premise** does not perform optimally within the Galactic plane, despite the high signal-to-noise. This is due to the large number of point sources present. The increase in masked data towards the centre of the Galactic plane results in fewer pixels of information from which to determine the thermal dust parameters. The worst estimations of τ_{353} appear within the Galactic plane. This is simply because the largest optical depth values occur within the Galactic plane and so the absolute differences between true and estimated values appear largest.

Similarly to Planck Collaboration et al. (2016c), the success of **premise** is proportional to the thermal dust signal-to-noise ratio; this is highlighted in Fig. 10 which shows the mean absolute percentage difference between true and **premise** estimated values across latitudes (the data are binned in intervals of 512). The peak in percentage difference within the Galactic plane (latitude 0°) is caused by a high fraction of the data being masked due to point sources. This peak can be seen to be larger for τ_{353} than for β and T . This is because the **premise** calculated τ_{353} loses thermal dust information due to point source masking twice. The τ_{353} **premise** maps are made using the total 857 GHz map after the masked data has been inpainted. The majority of this inpainting occurs within the Galactic plane and is an imperfect attempt to recover lost thermal dust information. This inpainted 857 GHz map is then used to determine the dust optical depth at 353 GHz using the **premise** determined temperature and spectral index values, which already suffer from a degradation in accuracy across the Galactic plane.

4.3.2 Thermal dust emission

The three **premise** full-sky parameter maps can be reconstituted to form the **premise** thermal dust estimates at 353, 545, 857 and 3000 GHz. We present these maps in the left-hand column of Fig. 11. The plots in the right-hand column show the median absolute flux density difference between the true and estimated dust maps when the data are binned at intervals of 512. We over-plot the nuisance flux term, CIB anisotropies (total CIB flux minus the constant CIB offset level) plus instrumental noise, in red. It can be seen that, for all frequencies, the largest discrepancies between true and estimated flux density occur within and approaching the Galactic plane. These discrepancies are far larger than the CIB and noise contamination simply because the thermal dust emission magnitude close to the Galactic plane is so high. On average the discrepancies are less than 10 per cent (for all frequencies): Table 1 states the percentage difference between true and **premise** estimated thermal dust emission for 50 per cent (medium), 68.3 per cent (1 σ), 95.4 per cent (2 σ) and 99.7 per cent (3 σ) of the full sky.

The 353 and 545 GHz latitude plots in Fig. 11 show absolute flux density differences at high latitudes at a lower level than the

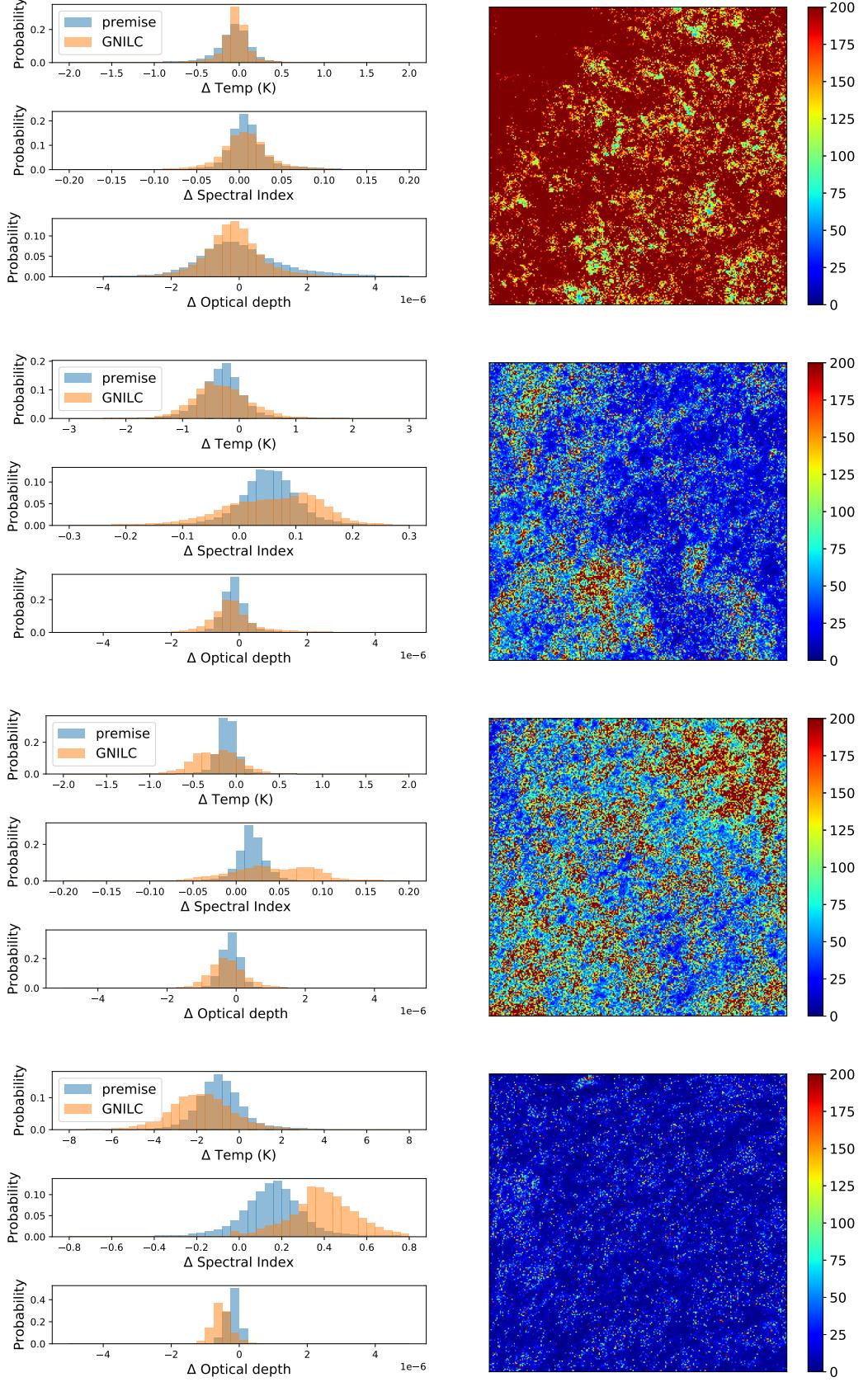


Figure 5. *Left:* Histograms of the differences between the calculated and true MBB parameters for the GNILC methodology and *premise*. *Right:* Signal-to-noise maps. Results for region 1 to 4 shown from top to bottom.

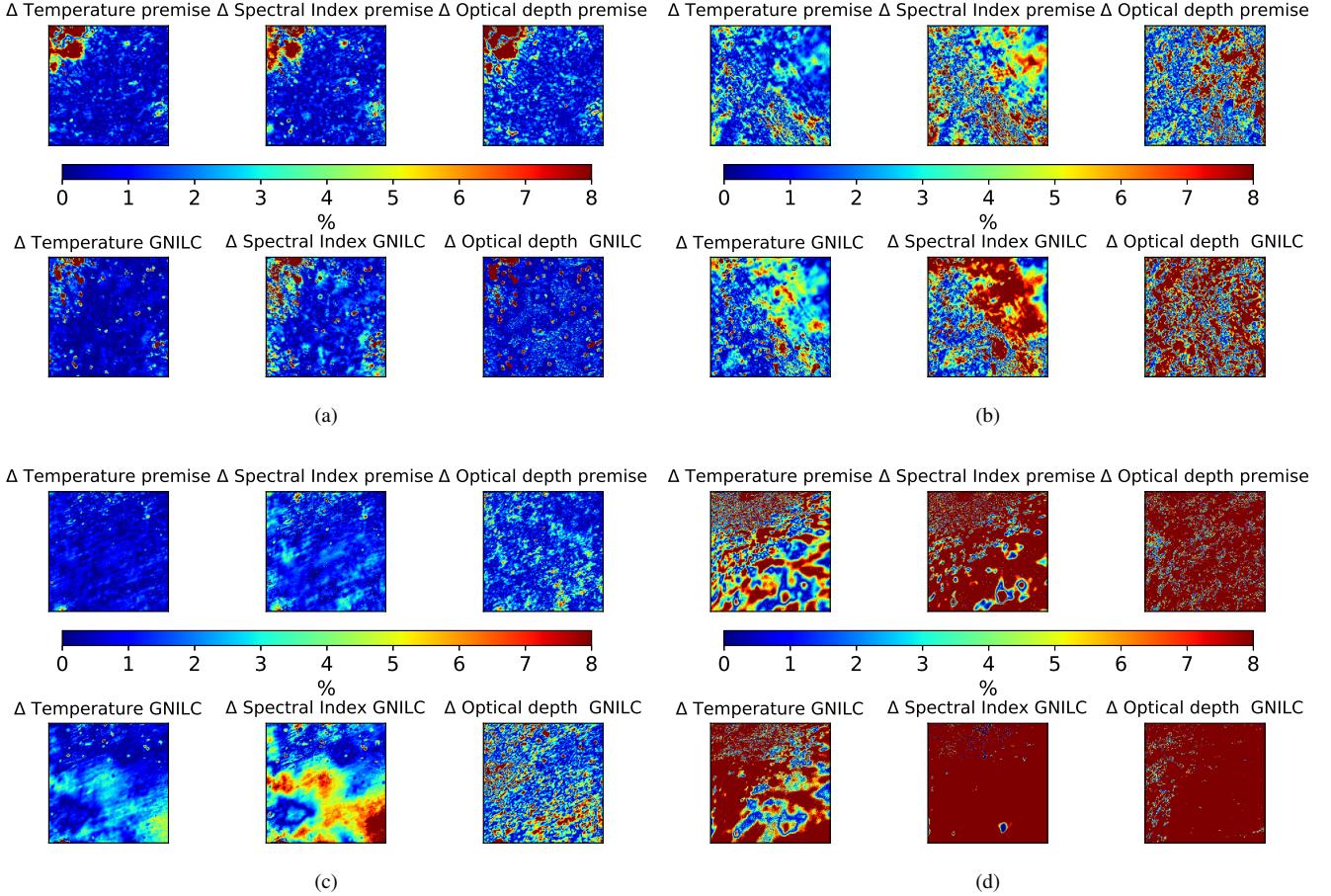


Figure 6. Maps of absolute percentage differences between the calculated and true MBB parameters for the GNILC methodology and `premise` for each pixel. Results for region 1 (a), 2 (b), 3 (c) and 4 (d).

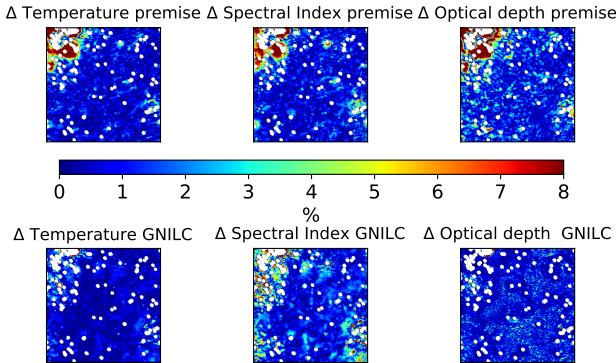


Figure 7. Region 1 map of absolute percentage differences between model and true temperature/spectral index for the GNILC methodology and `premise`.

actual 353 and 545 GHz CIB anisotropy level. The reason for this is that the 857 GHz total flux map has been used to provide the dust optical depth at 353 GHz and the CIB contamination at 857 GHz is far less dominant than at 545 or 353 GHz. This confirms that very little CIB contamination is present in the estimates of β and T ; the largest source of CIB contamination is from the optical depth. At 3000 GHz any residual CIB anisotropies plus noise present in

the optical depth parameter is extrapolated up to high frequencies following the MBB form and so increased in magnitude. Therefore the 3000 GHz latitude plot shows absolute flux density differences higher than the 3000 GHz CIB anisotropies plus noise level across all latitudes. This could be avoided if another, cleaner, thermal dust template was chosen - for example the 3000 GHz total flux map. The 857 GHz total flux map was chosen in this work as only the 353, 545 and 857 GHz simulated maps represent empirically taken data (the *Planck* HFI maps); the 3000 GHz total flux simulation represents the IRIS combination of *IRAS* and *COBE-DIRBE* maps.

5 CONCLUSIONS

Obtaining reliable thermal dust estimates from *Planck* HFI data is complicated by the presence of the CIB which ties the accuracy of any thermal dust estimates to the dust signal-to-noise ratio. We have presented a new method of parametric fitting based around informed, initial estimates of parameter values called `premise`. Full sky data are divided into patches of various size where each patch area contains only those pixels which share similar thermal dust properties. The MBB model is fit to each patch producing an all-sky estimate of the model parameters at a lower resolution than the raw data. These parameter estimates are then refined using a sparsity-based optimisation method to produce the final, full resolution parameter estimates.

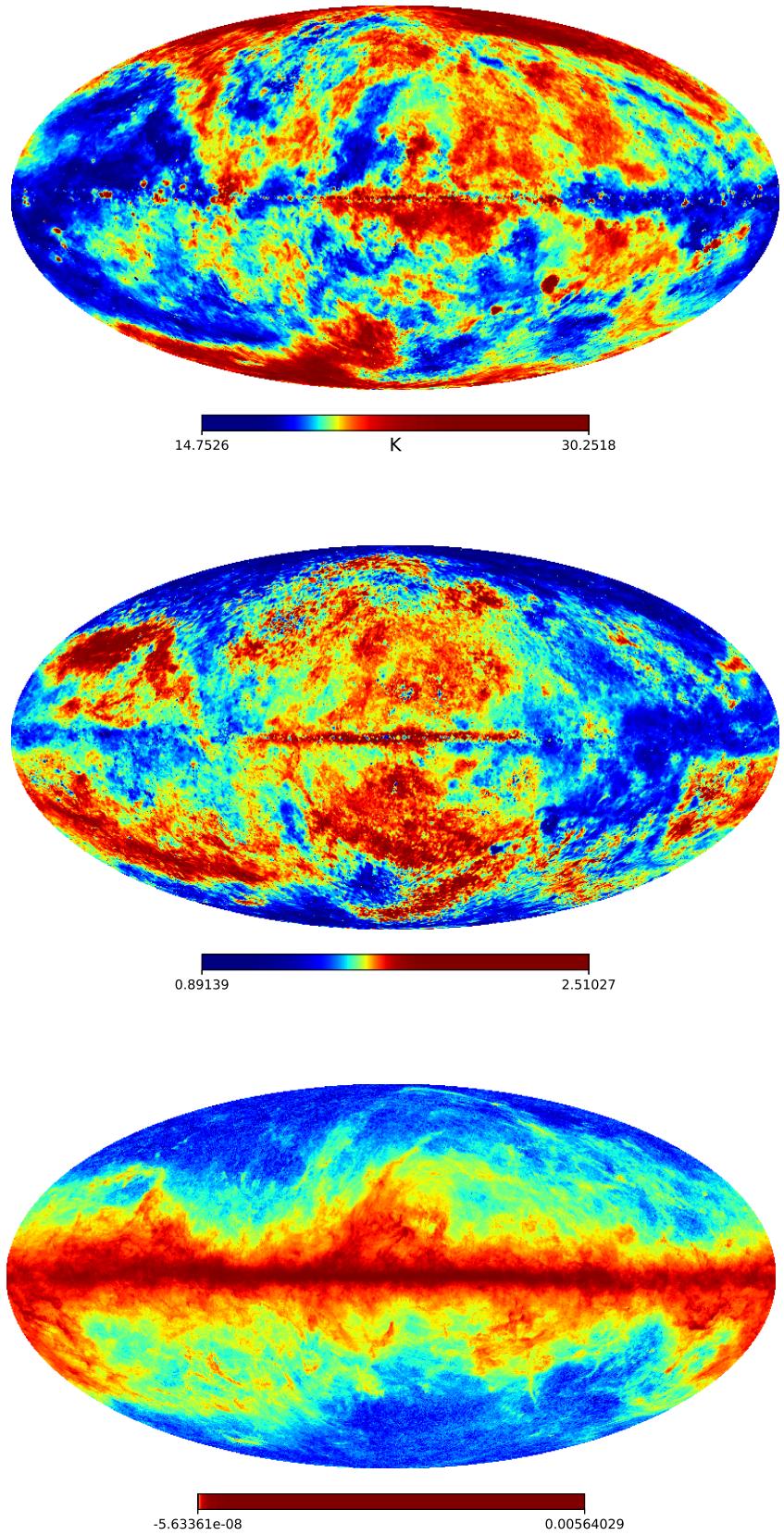


Figure 8. All-sky maps of thermal dust temperature (*top*), spectral index (*middle*) and optical depth at 353 GHz (*bottom*) produced by `premise`. The colour scale uses histogram equalisation.

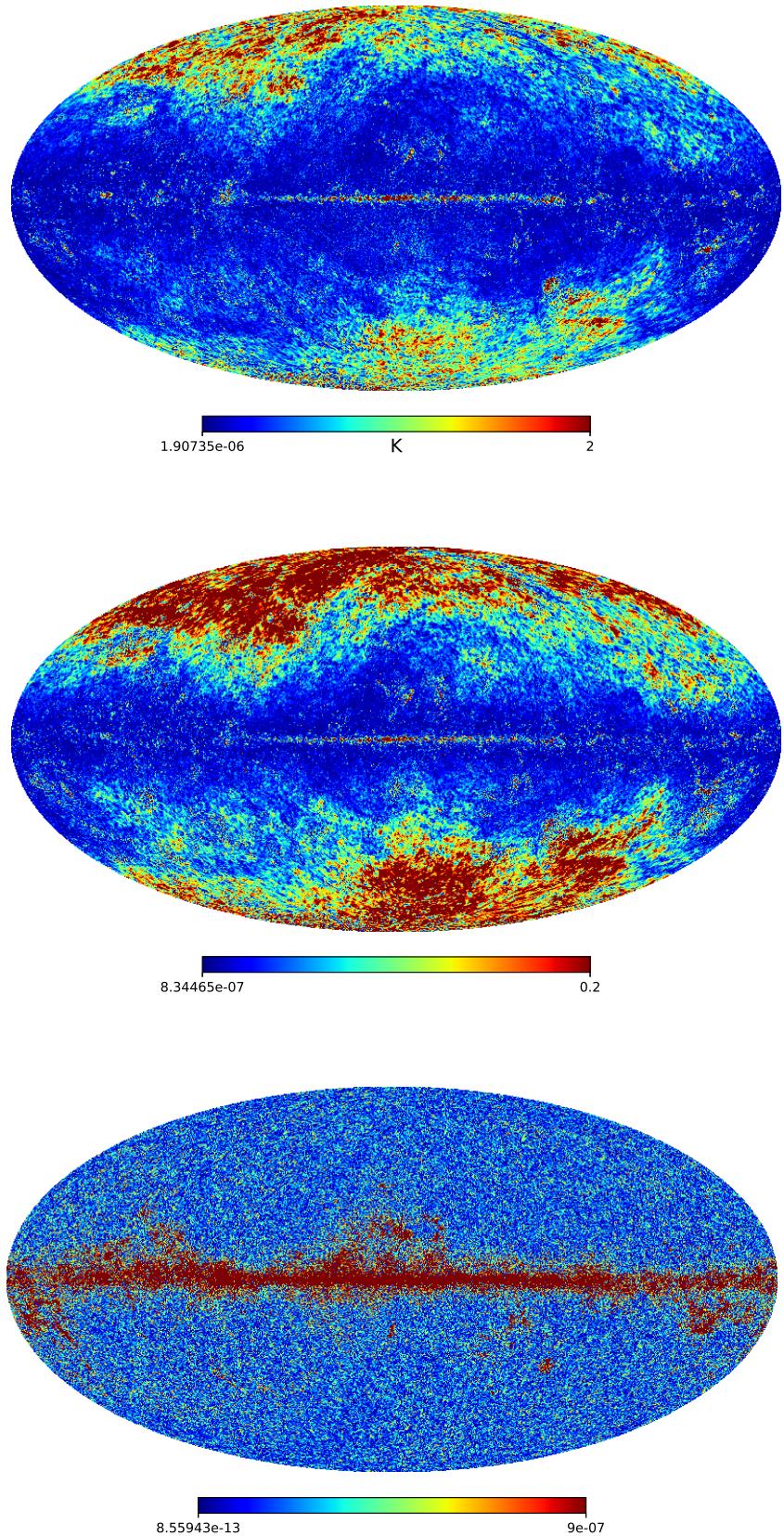


Figure 9. All-sky maps of the absolute differences between true and premise calculated values for thermal dust temperature (*top*), spectral index (*middle*) and optical depth at 353 GHz (*bottom*).

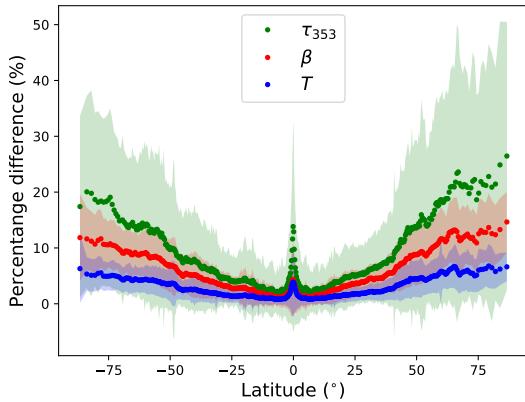


Figure 10. Mean absolute percentage difference between true and `premise` estimated parameter values as a function of Galactic latitude. Shaded area represent the standard deviation of the mean values.

By comparing `premise` to a GNILC-like method over select regions of the sky we find that the success of the GNILC-like methods is heavily dependent on the dust signal-to-noise ratio. `premise` can also be seen to suffer from this effect, however, by taking advantage of the sparse nature of thermal dust, `premise` demonstrates an increasing ability to outperform GNILC at parameter estimation as signal-to-noise worsens. This is because instead of increasing the level of smoothing to compensate for poor signal-to-noise, `premise` increases the number of wavelet coefficients that are prevented from contributing to the reconstructed thermal dust estimate. Therefore we can present MBB parameter maps at full resolution. Additionally, by splitting the method into a fast initial estimation and then a refinement step, we introduce a considerable time reduction in the computational time required for `premise` to perform on full sky, $N_{\text{side}} = 2048$ HEALPix maps.

In this work we fit for the thermal dust temperature, optical depth at 353 GHz and spectral index and produce all-sky maps of these three values with median absolute percentage deviations from the true simulation values of 1.7, 3.7 and 3.4 per cent, respectively. We use our dust parameters to reconstruct thermal dust estimates at 353, 545, 857 and 3000 GHz with median absolute percentage deviations from the true simulation values of 7.9, 7.3 and 2.2 and 7.4 per cent, respectively.

The main limitation of `premise` is accurate parameter estimation within the very centre of the Galactic plane (due to a loss of thermal dust information after point source masking).

ACKNOWLEDGEMENTS

REFERENCES

- Abrial P., Moudden Y., Starck J.-L., Afeyan B., Bobin J., Fadili J., Nguyen M., 2007, *Journal of Fourier Analysis and Applications*, 13, 729
- Beck A., Teboulle M., 2009, *SIAM Journal on Imaging Sciences*, 2, 183
- Bennett C. L., et al., 2013, *ApJS*, 208, 20
- Béthermin M., Wang L., Doré O., Lagache G., Sargent M., Daddi E., Cousin M., Aussel H., 2013, *A&A*, 557, A66
- Bobin J., Sureau F., Starck J.-L., 2016, *A&A*, 591, A50
- Compiègne M., et al., 2011, *A&A*, 525, A103
- Draine B. T., Li A., 2007, *ApJ*, 657, 810
- Finkbeiner D. P., Davis M., Schlegel D. J., 1999, *ApJ*, 524, 867
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
- Jones A. P., Fanciullo L., Köhler M., Verstraete L., Guillet V., Bocchio M., Ysard N., 2013, *A&A*, 558, A62
- Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, *A&A*, 440, 775
- Kowalski M., 2009, *Applied and Computational Harmonic Analysis*, 27, 303
- Lagache G., Puget J.-L., Dole H., 2005, *ARA&A*, 43, 727
- Mehta M. L., 2004, *Random Matrices*, Volume 142. Academic Press
- Meisner A. M., Finkbeiner D. P., 2015, *ApJ*, 798, 88
- Miville-Deschénes M.-A., Lagache G., 2005, *ApJS*, 157, 302
- Parikh N., Boyd S. P., et al., 2014, *Foundations and Trends in optimization*, 1, 127
- Planck Collaboration et al., 2014, *A&A*, 571, A11
- Planck Collaboration et al., 2016a, *A&A*, 594, A9
- Planck Collaboration et al., 2016b, *A&A*, 594, A12
- Planck Collaboration et al., 2016c, *A&A*, 596, A109
- Starck J.-L., Fadili J., Murtagh F., 2007, *IEEE Transactions on Image Processing*, 16, 297
- Starck J.-L., Murtagh F., Fadili J. M., 2010, *Sparse image and signal processing: wavelets, curvelets, morphological diversity*. Cambridge University Press
- Teschke G., Borries C., 2010, *Inverse Problems*, 26, 025007

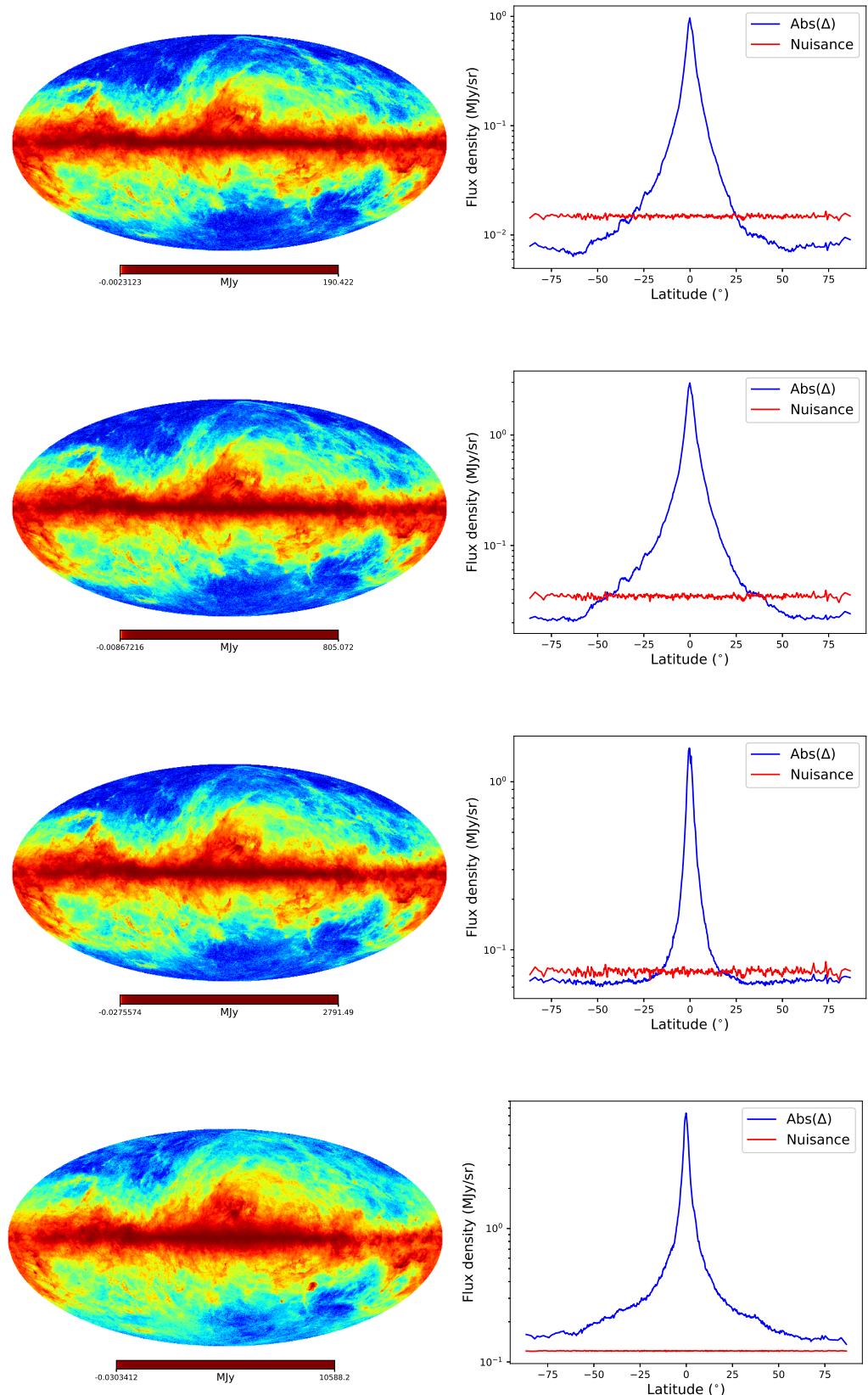


Figure 11. *Left column:* All-sky maps of thermal dust emission produced from the premise parameters. These maps have not been colour corrected and a histogram colour scale has been used. *Right column:* Plots of the absolute median differences between true and calculated thermal dust emission over latitude alongside the CIB anisotropies plus noise level (referred to as a ‘nuisance’ term). Results at frequencies 353, 545, 857 and 3000 GHz are shown from top to bottom.

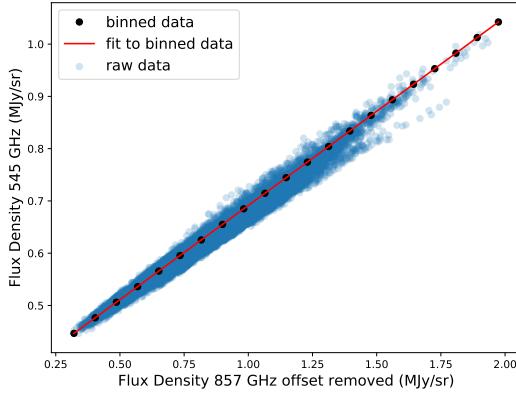


Figure A1. An example of the linear regression used to determine the large scale CIB offset for each frequency.

APPENDIX A: LARGE-SCALE CIB OFFSET

To determine the large-scale CIB offset for each frequency, data from the Leiden/Argentine/Bonn (LAB) H_I survey (Kalberla et al. 2005) were used. The LAB data are integrated over radial velocity so three maps are available: the low velocity map (-30 to 30 km s⁻¹), the intermediate velocity map (-100 to -30 km s⁻¹) and the high velocity map (-500 to -100 km s⁻¹).

The total flux (353, 545, 857 and 3000 GHz) and the LAB low and intermediate velocity all-sky maps were all downgraded to N_{side} 128 and smoothed to 1°. Two masks were then made which selected pixels for use depending on their column densities (N_{H_I}) within the low and intermediate velocity maps :

Mask	Low N _{H_I} (cm ⁻²)	Intermediate N _{H_I} (cm ⁻²)
1	< 2 × 10 ²⁰	< 0.1 × 10 ²⁰
2	< 3 × 10 ²⁰	–

Linear regression was used to determine the large-scale CIB offset at each frequency using the following masked data:

x axis data	y axis data	Mask applied
3000 GHz	low + intermediate LAB	2
857 GHz	low + intermediate LAB	1
545 GHz	857 GHz - offset ₈₅₇	1
353 GHz	857 GHz - offset ₈₅₇	1

Fig. A1 shows the linear regression between the 545 GHz and the 857 GHz total flux binned, mask 1 pixels. The 857 GHz offset has been removed from the 857 GHz data for the linear regression.

APPENDIX B: GNILC METHODOLOGY

The GNILC algorithm analysed in this work is based on that detailed in Planck Collaboration et al. (2016c). Our implementation is described here in full so that any differences may be highlighted. Instead of working on the sphere and using the needlet transform, the total flux maps were divided into twelve 2048X2048 pixel faces and the wavelet transform was used on the 2D faces. The optimum number of wavelet scales was empirically found to be six, instead of the ten scales used by Planck Collaboration et al. (2016c). This difference can be explained by the fact that Planck Collaboration

et al. (2016c) work on all 12X2048X2048 pixels at once and the number of wavelet scales required is proportional to the log of the number of data samples.

The following steps were completed for each wavelet scale (*j*):

1. The total flux maps were smoothed within the wavelet domain via convolution with a Gaussian of FWHM $2^j \times \frac{16\text{ pixels}}{2\sqrt{2}\log 2}$. This width may be different from that used by Planck Collaboration et al. (2016c).
2. The nuisance terms were collected as $\mathbf{N} = \text{CIB} + \text{CMB} + \text{instrumental noise}$. As simulation data were used the CIB, CMB and instrumental noise estimates were in fact the actual CIB, CMB and instrumental noise contributions to the total flux. The N_{obs} by N_{obs} nuisance covariance matrix was calculated as:
3. The covariance matrices of the smoothed total flux maps were calculated as $\mathbf{X}_{\nu_i} \times \mathbf{X}_{\nu_f}^T$ where $\mathbf{X}_{\nu_{i/f}}$ is the total flux within the wavelet domain at frequency *i/f*.
4. The N_{obs} by N_{obs} total flux covariance matrices for each smoothed pixel were whitened: $\mathbf{R}_{\text{nus}}^{-1/2} \mathbf{R}_{\text{tot}} \mathbf{R}_{\text{nus}}^{-1/2}$.
5. The eigenvectors of each whitened covariance matrix were ordered and the Akaike Information Criterion was used to select eigenvalues which deviated significantly from unity. Those eigenvectors (\mathbf{U}_s) gave the mixing matrix ($\mathbf{F} = \mathbf{R}_{\text{nus}}^{1/2} \mathbf{U}_s$) used to obtain the least-squares optimisation of thermal dust emission $\mathbf{F} (\mathbf{F}^T \mathbf{R}_{\text{tot}}^{-1} \mathbf{F}) \mathbf{F}^T \mathbf{R}_{\text{tot}}^{-1} \mathbf{X}$
6. If no eigenvectors were identified as significant then the total flux was believed to be dominated by the nuisance terms and so the signal contribution at that pixel and wavelet scale was masked out (set to zero).

The thermal dust contributions at each wavelet scale were recomposed to form the GNILC estimate of thermal dust emission for each frequency within pixel space.

This paper has been typeset from a \TeX / \LaTeX file prepared by the author.