Multi-resolution internal template cleaning: An application to the CMB polarization data

R. Fernández-Cobos, P. Vielva, R. B. Barreiro, E. Martínez-González
Instituto de Física de Cantabria, CSIC-UC, Avda. de los Castros s/n, E39005 Santander, Spain.

Abstract

We present a fast and robust method that uses a wavelet decomposition on the sphere to recover the CMB signal from microwave maps. This decomposition is implemented as a part of an internal template fitting that minimizes the variance of the resulting maps at each scale. We develop an application of the procedure on 7-year *WMAP* polarization data.

1 Introduction

The microwave data are actually a superposition of different contributions over the cosmic microwave background (CMB) radiation. This is the reason why an adequate component separation analysis becomes crucial to the estimation of the cosmological parameters encoded on the CMB fluctuations. The need is more acute in experiments dedicated to detect the B-mode polarization signal, where foreground amplitudes are clearly dominant. In other words, current and future ([1 − 7]) experiments are able to measure the CMB polarization anisotropies with such precision that foreground contamination will be the real limitation in data analysis.

The major galactic contaminants to the CMB polarization signal come from two physical processes: synchrotron emission and thermal dust. They are present at large scales and both are highly anisotropic, although their spatial variation is smooth. It is common to assume that their contribution can be factorized into two different terms: one of them depending only on the frequency and another depending on the position in the sky. In addition, extragalactic emissions also contribute to the signal. We refer, of course, to point sources and clusters that appear as compact objects. They are roughtly isotropically distributed and, in the case of extragalactic point sources, each object has a particular frequency dependence. Most of the component separation approaches assume that the brightest compact sources are masked or subtracted by other fitting approaches [8].

The generic goal of the component separation in this context is to recover the sky contributions. However, there is a whole zoology of methods that differ according to the

applications that are targeted. In this sense, we can talk about methods that recover only one component (tipically CMB signal, point sources, galaxy clusters...) or several ones at time, and the range of proposals includes Bayesian approaches, internal linear combinations (ILCs) and independent component analysis (see [9] for a review). As we are interested in recovering only the CMB signal, we focus on the first category. Specifically, our method can be situated in the context of the ILCs, in particular, a template cleaning whose coefficients are fitted in a wavelet decomposition that enables a multi-resolution analysis. This fitting by scales allows some effective variation of the coefficients in the sky. Therefore, our approach effectively lies in between standard template fittings (implemented in the real space) and more sophisticated parametric methods that assume physical features for the foreground emissions (e.g., $[10-12]$).

One of the advantages of choosing a template fitting approach is that many purposes require CMB maps at several frequencies. For instance, this allows a consistency check to verify whether any detected feature of the data is actually monochromatic or not (important in the context of non-Gaussianity analysis). Furthermore, the use of internal templates reduce any systematic effect that could be related to the calibration of the experiment. However, these templates are noisy, so we increase the total noise level when we remove them from the data. An alternative would be to add external templates, created from other independent observations or theoretical proposals. But the current knowledge of foreground emissions, in polarization, is not substantiated with a suitable ancillary data set.

2 The method

This method was presented in [13] and corresponds to the initial step of the map cleaning process in the SEVEM method ([14,15]).

We propose a CMB estimator as the subtraction between the original map at a given frequency ν and a linear combination of templates t_i:

$$\widehat{T}_{CMB}(\nu,p) = T(\nu,p) - \sum_{i=1}^{N_t} \beta_i^{\nu} t_i(p). \tag{1}$$

As the CMB signal expressed in units of thermodynamic temperature is the same in all the frequency range, an internal template is constructed as the difference of two maps corresponding to different frequency bands at the same resolution. The β_i coefficients are estimated by minimizing the variance of $\widehat{T}_{CMB}$ (note that we remove the monopole on all the maps):

$$E(\nu) = \sum_{p} \left[T(\nu,p) - \sum_{i=1}^{N_t} \beta_i^{\nu} t_i(p) \right]^2 . \tag{2}$$

2.1 *HEALPix* wavelet

We use the so-called *HEALPix* wavelet (HW,[16]), a discrete and orthogonal wavelet that provides a multi-scale decomposition on the sphere adapted to the *HEALPix* tessellation [17]. In this scheme, the resolution j of a map is a number such that $2^j = N_{side}$, where the N_{side} parameter is defined so that the number of pixels needed to cover the sphere

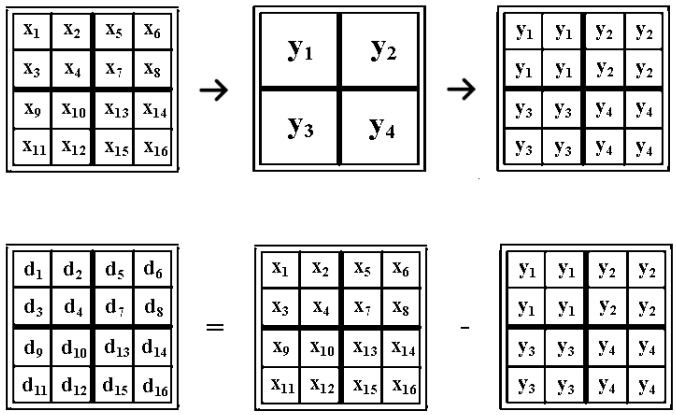

Figure 1: Outline of the construction of the detail coefficients at resolution j (d_i) by substracting the approximation coefficients at resolution j -1 (y_i) from the approximation coefficients at resolution j (x_i).

is $N = 12N_{side}^2$. Each map is decomposed in a series of maps from the resolution of the original map to the lowest resolution considered (j_{aprox}). These coefficient maps are called details, except the map at the lowest resolution. This last one is called the approximation and is constructed by degrading the original map to the appropiate resolution. As we show in figure 1, the detail coefficients are calculated at each resolution j by substracting the approximation coefficients at resolution j-1, udgraded to j, from the approximation coefficients at resolution j.

One of the advantages of this wavelet is that all the involved operations are linear, combinations of a small number of pixels, so its computational time is of the order of the number of pixels N_{pix}, whereas for other linear, but continuous wavelets, such as the spherical Mexican hat [18] or the needlets [19], this time is of the order of $N_{side}^{3/2}$.

We decompose both the map to be cleaned and the templates in the wavelet coefficient space. The minimization of the variance is then carried out at each scale, i.e., we minimize the following quantity:

$$E_j = \sum_p \left[T_j(p) - \sum_{i=1}^{N_t} \beta_{ij} t_{ij}(p) \right]^2 , \tag{3}$$

where the subscript j denotes the resolution.

Finally, we perform the synthesis of the wavelet coefficients to recover the CMB estimation in the pixel space.

3 Application: *WMAP* 7-year polarization data

One of the major complications in the application to *WMAP* polarization data is that noise is correlated [20]. The *WMAP* data are typically supplied at a *HEALPix* resolution of $N_{side} = 512$, but a more accurate version of the pixel-by-pixel correlation is only given at low resolution ($N_{side} = 16$). Because of that, we perform the cleaning in two cases: for low and high-resolution maps.

3.1 Low-resolution analysis

The *WMAP* data can be seen as a superposition of CMB signal, synchrotron emission, thermal dust and a contribution of instrumental noise. The *WMAP* team [21] used a template fitting implemented in the real space to clean the foreground emission in the Ka, Q, V and W bands using as templates the K band (to account for the synchrotron) and a low-resolution version of the model presented in [22] for the thermal dust, with polarization direction derived from starlight measurements.

However, we use only a synchrotron template computed as K-Ka. A previous analysis in the real space with another template that accounts to the thermal dust has showed that the contribution of the dust template is much smaller than the synchrotron one.

We clean the Q and U polarization components independently, applying the *WMAP* polarization analysis mask that excludes 26% of the sky.

The goodness of the procedure

As the CMB signal is clearly subdominant in the *WMAP* low-resolution data, we opted to compare our foreground-free maps with the expected level for a noisy sky, taking as reference the *WMAP* instrumental noise characteristics [1]. We construct a χ^2 distribution with a set of 10^4 simulations of the noise of our cleaned maps at the Q, V and W bands, $M_s(p)$, with $s \in \{1, ..., 10^4\}$. Namely, we calculate each value as

$$\chi^2_s = \sum_{p,q} M_s(p) N^{-1}_{Obs}(p,q) M^t_s(q), \tag{4}$$

where N_{Obs} is the noise correlation matrix. It is necessary a tipical number of the order of 10^6 simulations to estimate this matrix so that the distribution converges to the theoretical curve of a χ^2 with as many degrees of freedom as pixels outside the mask in the Q and U maps (in this case, we have 4518 degrees of freedom). This distribution characterizes the expected level of instrumental noise in our cleaned maps and we can associate the χ^2 value of the data with relative levels of the signal.

It is important to test our assumption that the CMB contribution is negligible in this context. We generated a set of 10^4 simulations with CMB and instrumental noise of the cleaned maps and we used them to compute another χ^2 distribution with the same matrix that we have already calculated with only the noise component. We have observed that the CMB provides a very small deviation between distributions (in particular, a shift of ~ 10 units of χ^2). Therefore, any significant deviation from the χ^2 noise distribution has to be assigned to foreground residuals.

We show a comparison between our results and those supplied by the *WMAP* team in figure 2. The values of the χ^2 for the case of the cleaned maps supplied by the *WMAP* team are calculated with the matrices that they provide. The χ^2 of the cleaned data maps is fully compatible with the instrumental noise level at the Q and V frequency bands. We even found a significant improvement at the Q band because the χ^2 value is shifted from 2σ to 0.5σ when our method is used. Furthermore, although we use a template that is noisier than those used by the *WMAP* team (K-Ka, vs K), the noise levels of our cleaned maps are lower. In particular, we have measured a difference of about 10% in the standard deviation of the data maps (this difference is also confirmed with simulations of instrumental noise).

[1] http://lambda.gsfc.nasa.gov/

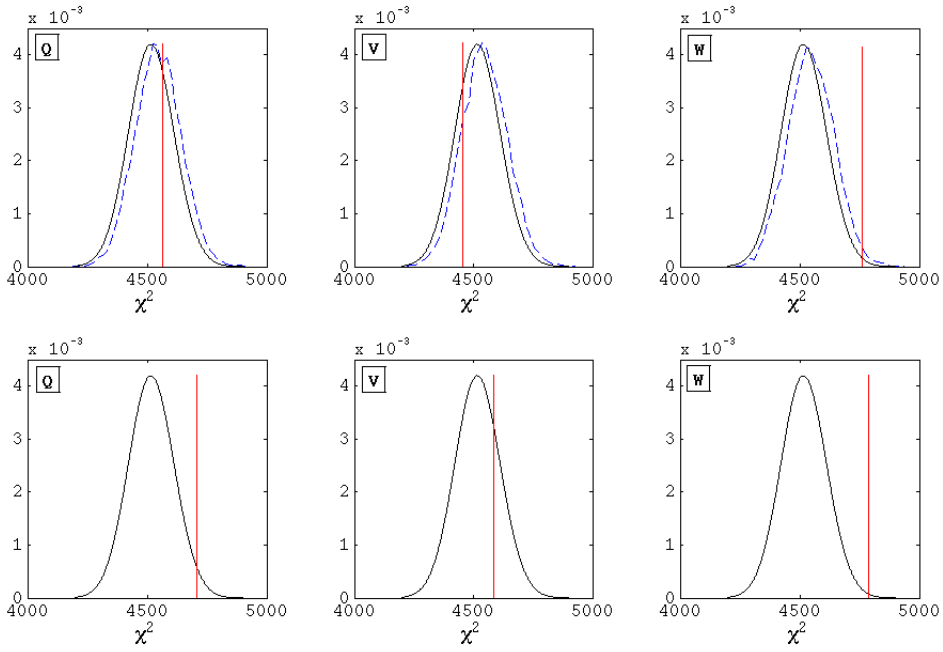

Figure 2: In the upper panels we show the χ^2 distributions of our cleaned maps, and in the bottom row we present corresponding plots for the *WMAP* team procedure. The solid line (in black) represents the theoretical curve of a χ^2 with as many degrees of freedom as pixels outside the mask (in other words, the effective number of pixels in Q and U maps: 4518 for this case). The dashed line (in blue) corresponds to the distribution calculated from simulations of our cleaned maps. The vertical line (in red) shows the χ^2 value of the foreground-free maps in each case. The columns refer to different frequency bands, from left to right: Q, V and W bands.

We notice that the χ^2 value of the cleaned data at the W band falls in the tail of the distribution. This deviation is even larger for the *WMAP* team procedure. The reason for this effect may be the presence of foreground residuals or a bad characterization of the instrumental noise in one of the involved channels. In this sense, and taking as starting point the full-sky covariance matrix of each DA, we computed analytically the noise covariance matrix of different combinations of the raw W-band DA maps. The χ^2 value for these maps (that contain only a combination of instrumental noise) shows that they are still compatible with the expected noise level. Nevertheless, it is significant that all the values are shifted to the left and that the most deviated combinations involve $W2$, followed by $W4$. We also performed the same analysis with the single-year foreground reduced maps supplied by the *WMAP* team for each DA at the W band and found that the χ^2 values are more deviated towards the tails for the W2 and W4 DAs.

Polarization power spectrum

We also estimate the polarization power spectrum using our cleaned maps of the $Q1$, $Q2$, $V1$ and $V2$ DAs. We calculate a pseudo cross-spectrum $\hat{D}_\ell^{AB}$ between any two differencing assemblies A and B as follows:

$$\hat{D}_\ell^{AB} = \sum_{\ell'} M_{\ell\ell'}^{AB} |p_{\ell'}|^2 B_{\ell'}^A B_{\ell'}^B \langle C_{\ell'}^{AB} \rangle + \langle N_\ell^{AB} \rangle, \tag{5}$$

where $A, B \in \{Q1, Q2, V1, V2 \mid A \neq B\}$; and, in the case of an EE power spectra,

$$\hat{C}_\ell^{AB} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} e_{\ell m}^A e_{\ell m}^{B*}, \tag{6}$$

where $e_{\ell m}$ are the spherical harmonic coefficients of the E-mode. We assume here a circular beam response, so the beam of the A map is denoted as B_ℓ^A, and the window function of the *HEALPix* pixel is given by p_ℓ. The bias introduced by the term of the noise cross-power spectrum, $\langle N_\ell^{AB} \rangle$, comes from the internal template fitting procedure, but it is small and controled by simulations. The coupling kernel matrix $M_{\ell\ell'}$ is described in [23] and, for the case of the polarization spectra, in Appendix A of [24]. The final angular power spectrum, $\hat{C}_\ell$, can be computed as a linear combination of the six different spectra weighted by the inverse of their variances:

$$\hat{C}_\ell = \left(\sum_i \frac{1}{\sigma_i^2} \right)^{-1} \sum_i \frac{1}{\sigma_i^2} \hat{C}_\ell^i, \tag{7}$$

where $i = AB$ and $\sigma_i^2 = \sigma_A \sigma_B$. These dispersions are supplied by the *WMAP* team in the *LAMBDA* web site.

We show the comparison of power spectra in figure 3. Both spectra are compatible with those that the *WMAP* team provides. On the one hand, most of the signal in C_ℓ^{EE} spectrum is at $\ell \lesssim 6$. On the other hand, the B-mode spectrum C_ℓ^{BB} signal is compatible with zero as expected.

The *WMAP* team uses a pixel-based likelihood to estimate the power spectra, whereas we use a combination of pseudo-spectra that is not optimal, so our error bars are larger than those obtain by them.

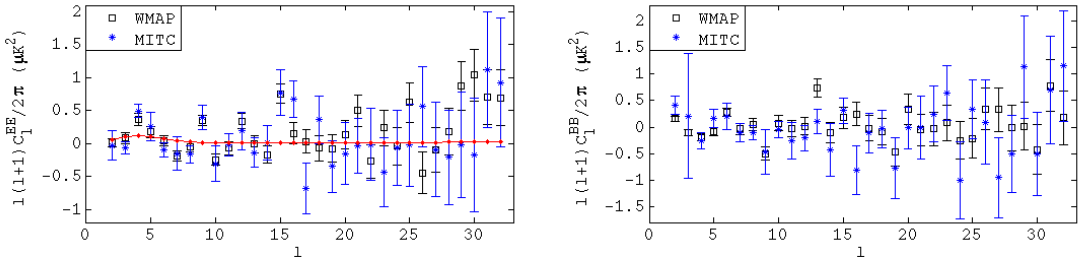

Figure 3: 7-yr *WMAP* polarization data: the left pannel corresponds to the polarization power spectrum EE and the right one with the power spectrum BB for low-resolution analysis. Squares (in black) represent the spectrum given by *WMAP* team and asterisks (in blue) represent our estimation (MITC, Multi-resolution Internal Template Cleaning). The solid line (in red) corresponds with the concordance ΛCDM model.

3.2 High-resolution analysis

The analysis of the *WMAP* data at $N_{side} = 512$ allows us to explore the effect of the cleaning at high multipoles where, a priori, the correlation of the noise is less important. In this regime, we only take into account the 2×2 noise covariance matrix of each pixel that relates the Q and U components. In this case, we employ two different internal templates: K-Ka to characterize the synchrotron radiation and $V1$-$W3$ to account for the thermal dust. We have chosen the V and W DAs with lower variance to be cleaned.

We estimate the auto power spectra following an analogous procedure to that exposed in the previous section. The error bars are estimated with 10^3 noise simulations (σ of the CMB component is negligible compared to the noise contribution). Bins are taken as a weighted average of the involved multipoles. We calculate these weights as the inverse of the variance of each C_ℓ.

The comparison of the auto power spectra is shown in figure 4. Our estimation is compatible with the spectra given by the *WMAP* team. The error bars associated to our approach are larger at high multipoles because, at this scales, the number of effective cross-spectra is much smaller than the number used by the *WMAP* team (where all the W-band DAs are available). For the first two points that account for the largest scales, we take the information of the low-resolution analysis described previously, since the correlation of the instrumental noise becomes important at these scales. As expected, the B-mode power spectrum is compatible with zero. In the E-mode spectrum, we can distinguish the acoustic peak around $\ell \sim 400$.

We also calculate an estimation for the cross-spectrum between temperature data and the E-mode polarization. The analysis is performed with the foreground reduced temperature maps supplied by the *WMAP* team and our cleaned maps of the Q- and U-Stokes parameters. In this case, the cross-spectra computed with the combination of two set of harmonic coefficients (corresponding to temperature and polarization) of the same DA is allowed. For the temperature maps, we use the temperature-analysis mask that the *WMAP* team supplies.

In this case, we take into account the CMB contribution to the cosmic variance in temperature in the estimation of the error bars. We use 10^3 simulations of CMB and instrumental noise, which undergo the same processes of cleaning and combination to obtain the cross-power spectrum TE, to estimate the error bars. The resulting cross-

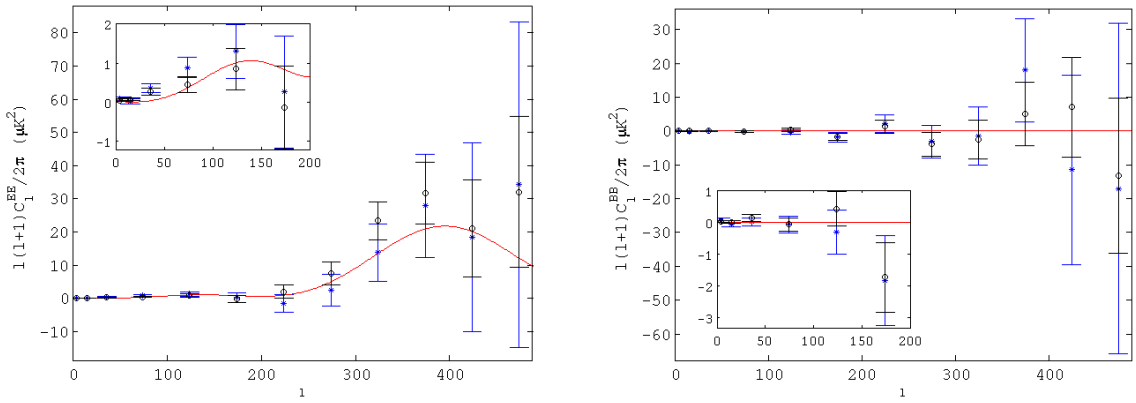

Figure 4: 7-yr *WMAP* polarization data: power spectrum EE (left pannel) and BB (right pannel) for high-resolution analysis. The asterisks represent our estimation and the circles are the spectrum supplied by the *WMAP* team. The solid line represents the fiducial model for C_ℓ^{EE} and the zero value for C_ℓ^{BB}.

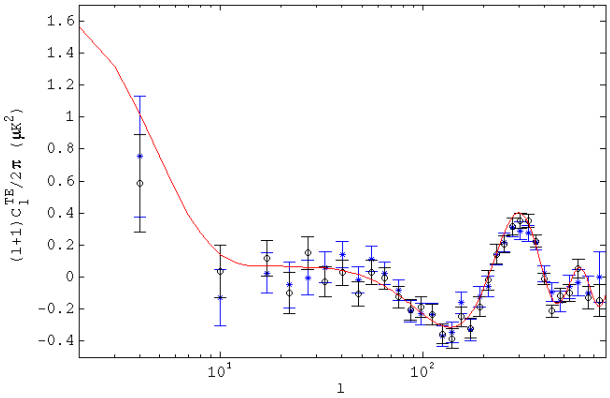

Figure 5: 7-yr *WMAP* data: cross-power spectrum TE. Asterisks represent our estimation and circles are the spectrum supplied by *WMAP* team. The solid line represents the fiducial model.

power spectrum is compatible with that given by the *WMAP* team and the comparison is presented in figure 5.

4 Conclusions

We present a cleaning procedure based on an internal template fitting method computed on the HW-coefficient space. The method allows to perform a multi-resolution analysis of the microwave maps, which can be interpreted as an effective variation of the spectral index in the sky. In addition, the compact support of the HW provides a good treatment of the incomplete sky coverage and the involved operations require less computational time than those that are necessary for the implementation of other continuous and isotropic wavelets.

We perform an application of the procedure: an analysis of the 7-yr *WMAP* polar-

ization data. We obtain satisfactory results cleaning both the high- and low-resolution *WMAP* data. The resulting maps enable us to compute power spectra, which seems to be compatible with those supplied by the *WMAP* team. At low resolution, we have evidence of a better cleaning of the Q-band map and, even when we are using noisier templates, the instrumental-noise levels of the final cleaned maps are lower than those of the maps provided by the *WMAP* team.

Acknowledgments

Authors acknowledge partial financial support from the Spanish *Ministerio de Economía y Competitividad* Projects AYA2010-21766-C03-01 and Consolider-Ingenio 2010 CSD2010-00064. RFC thanks financial support from Spanish CSIC for a *JAE-predoc* fellowship, co-financed by the ESF. PV thanks financial support from the Ramón y Cajal program. The authors acknowledge the computer resources, technical expertise and assistance provided by the *Spanish Supercomputing Network* (RES) node at Universidad de Cantabria. We acknowledge the use of *Legacy Archive for Microwave Background Data Analysis* (LAMBDA). The *HEALPix* package was used throughout the data analysis ([17]).

References

[1] Kogut A. et al., 2006, New Astron. Rev., 50, 1009
[2] Charlassier R. et al., 2008, preprint (arXiv:0805.4527v1)
[3] Grainger W. et al., 2008, Proc. SPIE, 7020, 68
[4] Rubiño-Martín J. A. et al., 2010, Highlights of Spanish Astrophysics V., p. 127
[5] Brown M. L. et al., 2009, ApJ, 705, 978
[6] Sievers J. L. et al., 2009, preprint (arXiv:0901.4540v2)
[7] Arnold K. et al., 2010, Proc. SPIE, 7741, 39
[8] Herranz D., Vielva P., 2010, IEEE Signal Processing Magazine, 27, 67
[9] Delabrouille J., Cardoso J. F., 2007, preprint (astro-ph/0702198v2)
[10] Eriksen H. K. et al., 2006, ApJ, 641, 665
[11] Eriksen H. K., Jewell J. B., Dickinson C., Banday A. J., Górski K. M., Lawrence C. R., 2008, ApJ, 676, 10
[12] Stivoli F., Grain J., Leach S. M., Tristram M., Baccigalupi C., Stompor R., 2010, MNRAS, 408, 2319
[13] Fernández-Cobos R., Vielva P., Barreiro R. B., Martínez-González E., 2012, MNRAS, 420, 2162
[14] Martínez-González E., Diego J. M., Vielva P., Silk J., 2003, MNRAS, 345, 1101
[15] Leach S. M. et al., 2008, A&A, 491, 597
[16] Casaponsa B., Barreiro R. B., Curto A., Martínez-González E., Vielva P., 2011, MNRAS, 411, 2019
[17] Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
[18] Martínez-González E., Gallegos J. E., Argüeso F., Cayón L., Sanz J. L., 2002, MNRAS, 336, 22
[19] Baldi P., Kerkyacharian G., Marinucci D., Picard D., 2009, Annals of Statistics, 37, 1150

[20] Jarosik N. et al., 2011, ApJS, 192, 14
[21] Gold B. et al., 2011, ApJS, 192, 15
[22] Finkbeiner D. P., Davis M., Schlegel D. J., 1999, ApJ, 524, 867
[23] Hivon E., Górski K. M., Netterfield C. B., Crill B. P., Prunet S., Hansen F., 2002, ApJ, 567, 2
[24] Kogut A. et al., 2003, ApJS, 148, 161