Non-iterative methods to estimate the in-flight noise properties of CMB detectors

P. Natoli[1], D. Marinucci[2], P. Cabella[1], G. de Gasperis[1], and N. Vittorio[1]

[1] Dipartimento di Fisica, Università di Roma "Tor Vergata", via della Ricerca Scientifica 1, 00133 Roma, Italy
[2] Dipartimento di Studi Geoeconomici e Statistici, Università di Roma "La Sapienza", via del Castro Laurenziano 9, 00161 Roma, Italy

Abstract. We present a new approach for statistical inference on noise properties of CMB anisotropy data. We consider a Maximum Likelihood parametric estimator to recover the full dependence structure of the noise process. We also consider a semiparametric procedure which is only sensitive to the low frequency behavior of the noise spectral density. Both approaches are statistically robust and computationally convenient in the case of long memory noise, even under nonstationary circumstances. We show that noise properties can be consistently derived by such procedures without resorting to currently used iterative noise-signal methods. More importantly, we show that optimal (GLS) CMB maps can be obtained from the observed timestream with the only knowledge of the noise memory parameter, the outcome of our estimators.

Key words. Cosmology: cosmic microwave background – methods: statistical, data analysis

1. Introduction

Recent high sensitivity, high resolution observations of the Cosmic Microwave Background (CMB) anisotropy are posing strong constraints on Cosmology (see, e.g., Balbi et al. 2000; de Bernardis et al. 2001). The ability to discriminate among different cosmological scenarios critically relies on the efficiency and accuracy of the data analysis pipeline, other than – obviously – on experimental know-how. In order to be efficient, data analysis has to deal with the size of the data sets, which is bound to increase with new space missions such as MAP[1] and PLANCK Surveyor[2]. In order to be accurate, the data analysis pipeline should implement optimal statistical techniques to reconstruct the correlation structure of receiver noise, i.e. the behavior of its spectral density. This is a critical point for data analysis of CMB anisotropy one-horned experiment. In fact, both Maximum Likelihood map making (see e.g. Wright 1996; Borrill 1999; Natoli et al. 2001) and angular power spectrum estimation (Tegmark 1997; Bond et al. 1998; Oh et al. 1999) heavily depend on the knowledge of the noise

spectral density, which of course must be determined from the data themselves.

Our paper is not the first to focus on estimating the noise properties out of a CMB anisotropy data stream. Ferreira & Jaffe (2000) advocated simultaneous estimation of signal and noise on the basis of a Bayesian approach. However, their estimates for the noise spectral density are piecewise constant with several discontinuities. Far from being realistic, this assumption does not apply to long range dependence models. Prunet et al. (2001) obtained a purely nonparametric estimate of the noise spectral density as a by-product of their iterative map making procedure. Even if cast in a frequentist framework, the resulting algorithm is very similar to Ferreira and Jaffe's proposal. It is worth noting that, in both cases, purely nonparametric estimates of noise properties lack a rigorous justification in the statistical literature if noise is a long memory process. In this case, all of these difficulties can be overcome by using well known and robust statistical techniques which are well supported by the most recent literature (Fox & Taqqu 1986; Dahlhaus 1995; Robinson 1995a,b; Giraitis & Taqqu 1999). In Sect. 2 we discuss these techniques in the context of CMB data analysis. In Sect. 3 we apply them to simulated datasets of the PLANCK and BOOOMERanG[3] experiments. Finally, in

---

[1] `http://map.gsfc.nasa.gov/`
[2] `http://astro.estec.esa.nl/SA-general/Projects/Planck/`

[3] `http://www.physics.ucsb.edu/~boomerang/`

Sect. 4 we draw our main conclusions and point out directions for future research.

2. Efficient estimation of the noise properties

We assume that the data, $\boldsymbol{d}$, consist of the sum of two uncorrelated time series:

$$d_t = s_t + n_t \tag{1}$$

for $t = 1, 2, \ldots, T$. Here $\boldsymbol{n}$ and $\boldsymbol{s}$ are noise and signal, respectively. The latter is modeled assuming linear dependence on the sky pattern:

$$s_t = \sum_{p=1}^{K} P_{tp} m_p. \tag{2}$$

We imagine the "map" $\boldsymbol{m}$ to be smeared by the instrumental beam and discretized into K elements (or pixels). The matrix $\boldsymbol{P}$ effectively defines the experimental scanning strategy by "unrolling" the sky map into a time series. We limit ourselves to the case of one-horned experiments and symmetric beam profile, for which the matrix $\boldsymbol{P}$ has a very simple structure, consisting of a single non zero entry (with value 1) per row. This entry corresponds to the pixel observed at a given time t. For very low signal to noise (S/N) ratio the timestream, $\boldsymbol{d}$, is noise dominated. In this section we pretend that this is the case. We will show below, in Sect. 3, how to drop this assumption.

2.1. Long memory processes

A process is said to be wide sense (or second order) stationary if its mean, variance and covariance functions are finite and constant over time: $\langle n_t \rangle = \mu$, $\langle n_t n_t \rangle = \gamma(0) < \infty$ and $\langle n_t n_{t-\tau} \rangle = \gamma(\tau)$, $\tau = \pm 1, \pm 2, \ldots$. For a noise process we shall take $\mu \equiv 0$. The Cramér Representation Theorem ensures that for any second order stationary process we can write:

$$n_t = \int_{-\pi}^{\pi} \exp(itf) \sqrt{P(f)} \, \mathrm{d}W(f), \tag{3}$$

where $W(f)$ is a random measure such that $\langle \mathrm{d}W(f) \rangle = 0$, while $\langle \mathrm{d}W(f) \mathrm{d}W(g) \rangle = \mathrm{d}f$ if $f = g$ and 0 otherwise (i.e. W represents a white noise measure). The function $P(f)$ is the spectral density function of the process $\boldsymbol{n}$, uniquely related to the sequence of covariances $\gamma(\tau)$ through the usual Fourier formulae:

$$P(f) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \exp(-i\tau f),$$

$$\gamma(\tau) = \int_{-\pi}^{\pi} \exp(-i\tau f) P(f) \mathrm{d}f.$$

Thus, wide sense stationarity implies:

$$\gamma(0) = \int_{-\pi}^{\pi} P(f) \mathrm{d}f < \infty,$$

that is, the spectral density needs always be integrable. State of the art microwave detectors produce "$1/f$" noise. We shall therefore focus on long memory (or long range dependent) noise. For this processes the spectral density is such that

$$\lim_{f \to 0} \frac{P(f)}{f^{-\alpha}} = G, \tag{4}$$

where $0 < G < \infty$ and $0 < \alpha < 1$. So, even if divergent at zero frequency, the spectral density is always integrable. Equation (3) can be discretized as follows:

$$n_t = \frac{1}{\sqrt{2\pi T}} \sum_{j=1}^{T} \exp(itf_j) \sqrt{P(f_j)} \, \varepsilon_j. \tag{5}$$

Here $f_j = 2\pi j/T$ are Fourier frequencies, while ε_j denotes a sequence of zero-mean, independent and identically distributed (i.i.d.) variables with unit variance. Equation (5) is widely used to generate random processes with a given spectral density $P(f)$ in Monte Carlo procedures. If we take, for instance, $P(f) = Gf^{-\alpha}$, straightforward algebra shows that:

$$\lim_{T \to \infty} \langle n_t n_t \rangle \simeq \lim_{T \to \infty} \frac{1}{2\pi T} \sum_{j=1}^{T} P(f_j) = \begin{cases} O(1), \ \alpha < 1 \\ O(\log T), \ \alpha = 1 \\ O(T^{\alpha-1}), \ \alpha > 1 \end{cases}.$$

Therefore, if $\alpha \geq 1$ the variance of $\boldsymbol{n}$ grows with time, i.e. the process is not stationary. We shall first focus on the stationary case $\alpha < 1$. This assumption will be relaxed to $\alpha < 2$ in Sect. 2.4.

2.2. Fully parametric procedures

Assume first that the functional form of the spectral density for the process $\boldsymbol{n}$ is known and completely specified by a set of parameters $\theta \equiv (\theta_1, \theta_2, \ldots, \theta_N)$. For "$1/f$" noise it is customary to write:

$$P(f, \theta) = A \left\{ 1 + \left(\frac{f_k}{f} \right)^{\alpha} \right\} \tag{6}$$

$\theta = (A, \alpha, f_k)$, where A is an amplitude, f_k is the so-called knee frequency and α is the spectral index introduced above. Throughout this paper we shall assume that $\boldsymbol{n}$ is Gaussian even if this condition is not essential for our conclusions to hold. Under Gaussianity $\boldsymbol{n}$ has a multivariate density function:

$$\Psi(\boldsymbol{n}) = \frac{1}{\sqrt{(2\pi)^T |\boldsymbol{N}|}} \exp \left\{ -\frac{1}{2} \boldsymbol{n}' \boldsymbol{N}^{-1} \boldsymbol{n} \right\}, \tag{7}$$

where $\boldsymbol{N}$ is the noise covariance matrix and the prime $(')$ denotes transposition. Now, for large T the matrix $\boldsymbol{N}^{-1}$ is nearly circulant and can be approximated as:

$$\boldsymbol{N}^{-1} \simeq Q^{\dagger} \Lambda^{-1} Q \tag{8}$$

where $\Lambda = \mathrm{diag}\{P(f_1), \ldots P(f_{T/2})\}$, the dagger $(\dagger)$ denotes complex conjugation coupled with transposition and Q is the rectangular $(T/2 \times T)$ matrix

$$Q_{jt} = \frac{\mathrm{e}^{itf_j}}{\sqrt{2\pi T}}. \tag{9}$$

The sample Discrete Fourier Transform (DFT)

$$Q\,\boldsymbol{n} \equiv \{w_T(f_1), \ldots, w_T(f_{T/2})\}\,,$$

has elements

$$w_T(f_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} n_t e^{itf_j}\,.$$

The sample periodogram is defined as

$$I_T(f_j) = w_T(f_j) w_T(f_j)^*\,,$$

where the asterisk denotes conjugation.

From Eq. (7), it is immediate to derive the log-likelihood function for θ as

$$\mathcal{L}(\theta; \boldsymbol{n}) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{N}(\theta)| - \frac{1}{2}\boldsymbol{n}'\boldsymbol{N}(\theta)^{-1}\boldsymbol{n}\,.$$

Maximization of this form in the time domain is a computationally unfeasible task even for moderately large values of T, let aside $T \gg 10^7$ that will become customary in CMB experiments. Based only on the approximations in Eq. (8) and building upon ideas that trace back to work by Whittle (1953), Fox & Taqqu (1986) have suggested the following frequency domain expression, valid up to a constant term:

$$\begin{aligned}
-2\mathcal{L}(\theta; \boldsymbol{n}) &\simeq \log|Q\Lambda Q^\dagger| + \boldsymbol{n}'Q^\dagger\Lambda^{-1}Q\boldsymbol{n} \\
&= \log|Q^\dagger Q| + \log|\Lambda| + (Q\boldsymbol{n})^\dagger\Lambda^{-1}Q\boldsymbol{n} \\
&= \sum_{j=1}^{T/2}\log P(f_j;\theta) + \sum_{j=1}^{T/2}\frac{I_T(f_j)}{P(f_j;\theta)}. \quad (10)
\end{aligned}$$

The Whittle Estimate (WE) for the noise parameters, θ, is then defined as:

$$\widehat{\theta} = \arg\min_{\theta}\left\{-\sum_{j=1}^{T/2}\log P(f_j;\theta) - \sum_{j=1}^{T/2}\frac{I_T(f_j)}{P(f_j;\theta)}\right\},$$

where $\arg\min_x\{f(x)\}$ denotes the value x which minimizes $f(\cdot)$. The same authors show that, under regularity conditions, these estimates are $\sqrt{T}$–consistent, asymptotically unbiased and Gaussian, meaning that

$$\lim_{T\to\infty}\sqrt{T}\,(\widehat{\theta} - \theta) \sim \mathcal{N}(0, \Sigma) \quad (11)$$

where $\mathcal{N}$ denotes a Gaussian multivariate with covariance matrix Σ whose explicit expression is given by Fox and Taqqu (1986). The estimate $\widehat{\theta}$ is computationally very convenient and Dahlhaus (1989) proves that it is asymptotically equivalent to the corresponding estimate in the time domain, i.e. to standard Maximum Likelihood. It follows immediately that the WE is absolutely efficient in the Cramér-Rao sense, that is it achieves minimum variance. Generalizations to a non Gaussian framework are considered by Giraitis & Surgailis (1990) and Giraitis & Taqqu (1999). Thus, provided that the functional form of the spectral density is known a priori, WE solves the problem of optimal inference of the noise properties given that the observed timestream is noise dominated.

2.3. Semiparametric procedures

In some situations the functional form of the spectral density may not be known. Nonetheless, in the presence of long range dependence, knowledge of the low frequency behavior of the noise spectral density is all that is needed to implement optimal filtering procedures (Dahlhaus 1995). Thus, determining α is sufficient for Generalized Least Squares (GLS) or Maximum Likelihood (ML) map making. This fact calls for statistical methods which, rather than parameterize the full spectral density, only rely on the much milder condition of Eq. (4). One such procedure, the so-called Log Periodogram Estimate (LPE), was introduced by Geweke and Porter-Hudak (1983) and discussed rigorously by Robinson (1995a). Consider the identity:

$$\log I_T(f_j) = \log G - \alpha\log f_j + \log\frac{P(f_j)}{Gf_j^{-\alpha}} + \log\frac{I_T(f_j)}{P(f_j)}\,, (12)$$

where $j = 1, 2, \ldots, m < T/2$ and $m = m(T)$ is chosen such that $\lim_{T\to\infty} 1/m = 0$ and $\lim_{T\to\infty} m/T = 0$. Under this Asymptotic Bandwidth Condition (ABC) and in view of Eq. (4), one has that $\log[P(f_j)/Gf_j^{-\alpha}] \to 0$ whereas $\log[I_T(f_j)/P(f_j)]$ can be approximated as a sequence of zero-mean and nearly uncorrelated residuals, with finite variance. Therefore, it seems natural to consider Eq. (12) as a regression model. This heuristic is made rigorous by Robinson (1995a), where the Ordinary Least Square (OLS) estimate is considered:

$$\widehat{\alpha} = \arg\min_{\alpha, G}\sum_{j=1}^{m}\left\{\log I_T(f_j) - G - \alpha\log f_j\right\}^2\,. \quad (13)$$

The intuitive meaning of the ABC is that we only consider a vanishing (as $T \to \infty$) subset of frequencies around the origin. Because of this, we are actually discarding most of the available information as it is the case for semiparametric procedures. The practical consequence, highlighted in Robinson (1995a), is that $\widehat{\alpha}$, despite being asymptotically Gaussian and unbiased, is only $\sqrt{m}$-consistent (in the sense of Eq. (11)). It has therefore asymptotical efficiency zero with respect to WE when the spectral density is correctly parameterized. In principle, a misspecified model may entail inconsistent estimates of all parameters, and in particular of α, which is the parameter of interest for many, if not most, applications. In this context, robust semiparametric procedures may represent a more reliable choice.

Robinson (1995b) considers another semiparametric method which can be viewed as a narrow-band analogous of WE. Its properties and motivating rationale, however, are too close to LPE to warrant independent consideration in this paper.

A somewhat intermediate attitude between the fully parametric WE and the semiparametric LPE has been recently set forth by Moulines & Soulier (1999). The idea is basically to use Eq. (4) to analyze the lowest frequencies, where the long range dependent behavior shows up, and to consider a series expansion into orthogonal components for

the remaining part of the spectral density. These estimates are in principle appealing because they can be shown to converge as fast as $\sqrt{T/\log T}$, provided that the spectral density $P(f)$ is sufficiently well behaved close to the origin (see Moulines & Soulier 1999 for more details). However, they are computationally more costly than WE and LPE because they require multiple regression with hundreds or thousands of estimated explanatory terms.

2.4. The nonstationary case

The analysis in the previous section assumes stationarity of the noise series. This assumption may turn out to be too strong over a long time span. For instance, the variance of the noise might grow (slowly) as the observation goes on. A general model for such a process might be

$$n_t = \sum_{j=0}^{t} \psi_j \, \varepsilon_{t-j},$$

where ε_t represents a sequence of independent innovations with zero mean and finite variance, whereas the weight sequence $\psi_j \simeq \kappa j^{\alpha/2-1}$, $(\kappa > 0)$. For $\alpha < 1$, it is not difficult to see that $\boldsymbol{n}$ is asymptotically stationary and satisfies Eq. (4) as $t \to \infty$. For $1 < \alpha < 2$ Velasco (1999) proves that, although the spectral density is no longer defined, the expected value of the periodogram maintains the same behavior as in Eq. (4). In the case $\alpha = 2$, $\boldsymbol{n}$ behaves (asymptotically) as a pure random walk process:

$$n_t \simeq \kappa \sum_{j=0}^{t} \varepsilon_j.$$

The case $\alpha > 2$ can also be considered. Robinson & Marinucci (2001) investigate the behavior of the periodogram for any positive value of α. However, this would entail a superlinear growth of the variance with time. We rule this out as experimentally questionable.

LPE and WE are shown to yield consistent estimates even under the condition $\alpha < 2$ (Velasco 1999; Robinson & Velasco 2000). This fact allows us to also consider values of α greater than unity in the simulations we will carry out in the next section. As a consequence, the procedures advocated in this paper enjoy a marked advantage with respect to those so far considered in the CMB literature, to the extent in which the latter analyses are deeply rooted in a stationary framework and apparently lack rigorous justification otherwise.

3. Numerical applications

The methods outlined above assume that the noise time series, $\boldsymbol{n}$, is known. Obviously this is not the case for a real world experiment where the observed sample is a combination of noise and signal. This is precisely the reason why iterative methods have grown attention in the literature. In this paper, however, we will follow a different – yet simpler – approach. That is, we will test the efficiency of the statistical methods outlined above on the simplest noise estimator, obtained by subtracting the OLS (i.e. naively coadded) map from the data. To be more specific, let us write Eq. (1) in matrix form as:

$$\boldsymbol{d} = \boldsymbol{P}\boldsymbol{m} + \boldsymbol{n}.$$

The OLS estimator is then:

$$\tilde{\boldsymbol{n}} \equiv \boldsymbol{d} - \boldsymbol{P}(\boldsymbol{P}^T\boldsymbol{P})^{-1}\boldsymbol{P}^T\boldsymbol{d} = [\boldsymbol{I} - \boldsymbol{P}(\boldsymbol{P}^T\boldsymbol{P})^{-1}\boldsymbol{P}^T]\boldsymbol{n} \quad (14)$$

which does not include any contribution from the signal, independently of the specific S/N ratio of the timestream (Natoli et al. 2001).

The plan of this section is as follows. First, we test the efficiency of WE and LPE on the pure noise $(\boldsymbol{n})$ timestream. We then compare these results with those obtained by using the $\tilde{\boldsymbol{n}}$ estimator. To assess this point, we make Monte Carlo simulations for the PLANCK mission. We then use our noise estimates to perform map making on simulated datasets for both PLANCK and BOOMERanG. In doing so, we also compare the outcome of our noise estimation method with our implementation of the iterative scheme proposed by Prunet et al. (2001) for BOOMERanG.

3.1. LPE and WE efficiency on n and ñ

Our first test bed is the 30 GHz channel of PLANCK/LFI. We remind here that PLANCK spins at 1 rpm, has a boresight angle of $85°$ and observes the same circle of the sky for 1 hour (every hour the spin axis is moved, say along the ecliptic, by $2.5'$). The mission is planned to remain in operation for at least 14 months, resulting in approximately two nearly full sky coverages. This corresponds to $\sim 10^9$ samples for each of the two LFI 30 GHz horns. The size of this timestream makes Monte Carlo simulations quite prohibitive. Therefore, we have decided to use only $\sim 2^{24}$ ($\sim 2 \times 10^7$) observations corresponding to ~ 140 rings (each ring is scanned 60 times) and covering two slices, each about $12°$ degrees wide. This means that we process ~ 12 days of observation. For the moment, we choose the usual "$1/f$" noise given in Eq. (6). As discussed in Sect. 2, we will examine both the $\alpha < 1$ and the $\alpha \geq 1$ (i.e., non stationary) case. The values of A and f_k are chosen accordingly to the PLANCK specifications: $A \simeq 4 \times 10^4 \, (\mu K\sqrt{s})^2$ and $f_k = 0.1$ Hz. The simulations are performed by first generating a pure noise timestream, $\boldsymbol{n}$, according to Eq. (5) with the power spectrum given in Eq. (6). This is accomplished by use of standard FFT techniques. The $\tilde{\boldsymbol{n}}$ estimator is trivially computed using Eq. (14). The periodograms of the series are then estimated by binning together a given number of neighboring frequencies (in the statistical literature this procedure is known as "pooling of the periodogram"). This choice has two advantages: (1) it brings down the size of the dataset (from 2^{12} proper frequencies to, say, a few thousands) while preserving most, if not all, of the relevant information; (2) it is potentially beneficial from
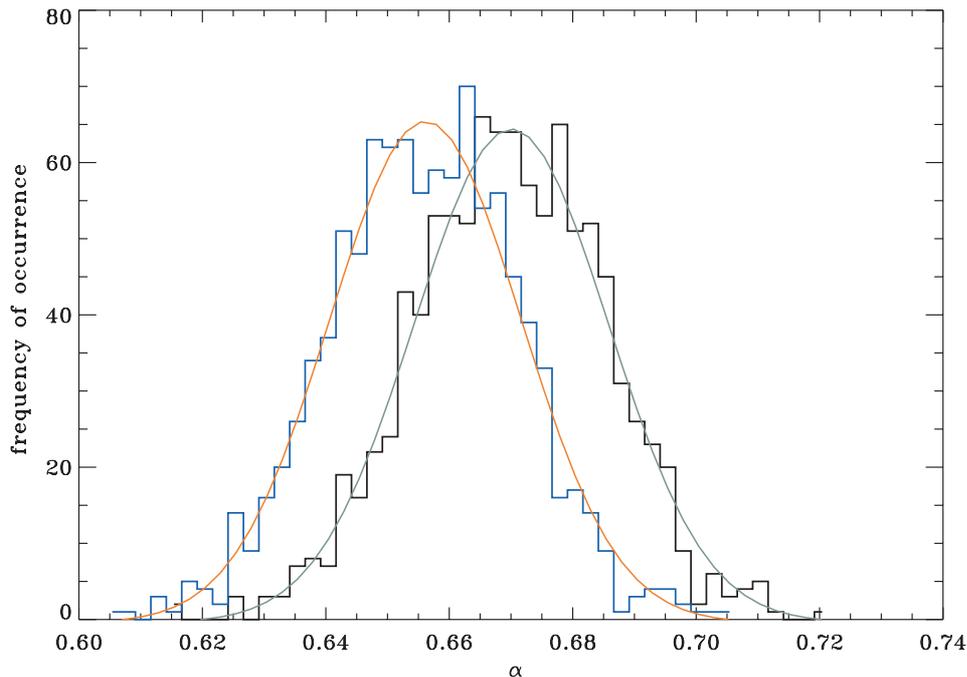
Fig. 1. LP estimates for $\alpha = 0.7$ on $\boldsymbol{n}$ (black curve) and $\tilde{\boldsymbol{n}}$ (dark blue curve). The bell shaped curves overplotted are Gaussian fits to the data.

the point of view of the efficiency of the estimator (see Robinson 1995a for discussion concerning the LPE case). Simulations show that WE and LPE are affected, though not dramatically, by the number of frequencies pooled together; in fact, if we bin too many frequencies we can even degrade the quality of the estimator. We found that a reasonable choice is to pool together 64 frequencies. As discussed in Sect. 2, LPE only exploits the m lowest frequencies of the periodogram. In our simulations, we find that an optimal number is $m \simeq 80$[4]. While the implementation of LPE is straightforward, requiring only a linear regression routine, the same is not true for WE. In fact, in the latter case we have to minimize Eq. (10) over the chosen parameter space which is spanned by all physically acceptable values of A, f_k and α. A minimization routine is then needed. A seemingly good candidate is a direction set algorithm (see e.g. Powell's method described in Press et al. 1992) which is unfortunately too slow for our purposes. A much faster choice is to use a variable metric method, such as the BFGS algorithm (see again Press et al. 1992). This method requires that the gradient of the function of interest is calculated at an arbitrary point, an information readily obtained by differentiating Eq. (6) w.r.t. the parameters.

We have performed 1000 realizations for $\alpha = 0.7$, 1 and 1.3, respectively. Our results for LPE and WE are shown as histograms in Figs. 1 to 6 and summarized

in Table 1 where mean values and standard deviations (across realizations) are reported.

Our first conclusion is that both LPE and WE are not significantly degraded when $\tilde{\boldsymbol{n}}$ is considered. In fact, the bias is at most 1% for WE and only slightly larger for LPE. The results for standard deviations are also comparable. Indeed the relative performance of LPE increases as α grows. In general, WE performs slightly better, as expected. However, the simple, robust LPE stands up as a very convincing alternative in the absence of precise a priori information on the functional form of the spectral density. This point will be further discussed below.
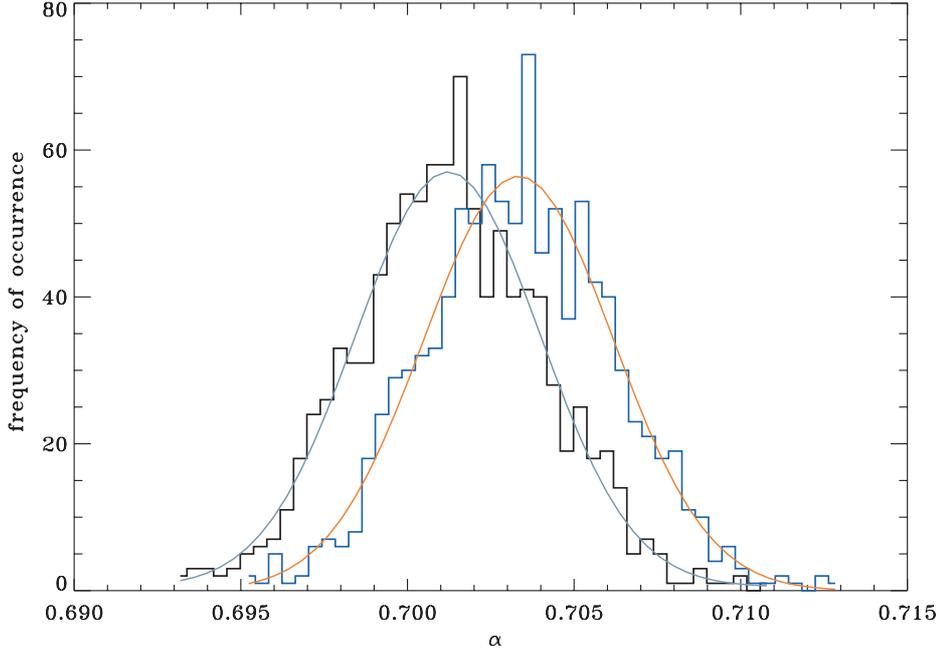
We have also implemented the Moulines & Soulier (1999) broad-band log periodogram estimator mentioned in Sect. 2. The resulting estimates, however, were extremely close to those obtained with LPE and we thus decided to omit their presentation for the sake of brevity.

3.2. Effects on map making

In order to understand how WE and LPE impact on map making we evaluated GLS maps out of 50 data streams simulated as discussed in Sect. 3.1. We recall here that the GLS estimate for the map is:

$$\tilde{\boldsymbol{m}} = (\boldsymbol{P}'\boldsymbol{N}^{-1}\boldsymbol{P})^{-1}\,\boldsymbol{P}'\boldsymbol{N}^{-1}\boldsymbol{d}\,,$$

where $\boldsymbol{N}$ is the noise covariance matrix introduced in Eq. (7). The Monte Carlo pipeline relies on the following scheme: (1) generation of the noise timestream, $\boldsymbol{n}$; (2) computation of $\tilde{\boldsymbol{n}}$; (3) derivation of WE and LPE. The maps are computed using the iterative map making algorithm discussed in Natoli et al. (2001).

---

[4] In the statistical literature it is occasionally suggested to trim (i.e. drop) the lowest ($\sim$10) frequencies of the periodogram. We have verified that, in our case, this choice has no significant impact on results.

Fig. 2. Same as Fig. 1 but for WE. Note the shrinking in the horizontal scale.

Table 1. Recovery of relevant parameters for WE and LPE.

α_{input}	0.7		1.0		1.3	
	n	$\tilde{n}$	n	$\tilde{n}$	n	$\tilde{n}$
LPE $\widehat{\alpha}_{\text{LP}}$	0.671 ±0.015	0.657 ±0.016	1.013 ±0.017	0.988 ±0.016	1.361 ±0.021	1.314 ±0.016
WE $\widehat{\alpha}_{\text{W}}$	0.7015 ±0.0028	0.7036 ±0.0029	1.0042 ±0.0033	1.0082 ±0.0039	1.3223 ±0.0054	1.317 ±0.030

We note that WE return estimates of A, f_k and α. Only the latter two are required to build $\mathbf{N}$, the scale factor being irrelevant. On the other hand, LPE only returns estimates of α. So, a fiducial value for f_k is used to compute $\mathbf{N}$. This approach finds its theoretical justification in Dahlhaus (1995). He shows that taking

$$N_{jk} \propto \int_{-\pi}^{\pi} \mathrm{e}^{i(j-k)f} f^{-\alpha} \mathrm{d}f \qquad (15)$$

provides estimates asymptotically equivalent to exact GLS under the only assumption that Eq. (4) holds. Note that Eq. (15) only depends on α.

A natural figure of merit to assess the relative performance of LPE and WE is the angular power spectrum of the maps. The latter is computed as

$$C_\ell = \frac{1}{2\ell+1} \sum_{m=\ell}^{\ell} |a_{\ell m}|^2 \,, \qquad (16)$$

where the $a_{\ell m}$'s are coefficients in the spherical harmonic expansion of the noise maps. We are not interested here in a rigorous power spectrum estimation but, rather, in a comparison between LPE and WE. This justifies the use of Eq. (16) which neglects corrections for, e.g., the sky

coverage and detailed shape of the surveyed region. The results are shown in Fig. 7 (note that the C_ℓ's are averaged over the simulation index). The curves for WE and LPE are virtually indistinguishable. This confirms that only the knowledge of α is relevant for GLS map making. Also, the small differences between LPE and WE become de facto immaterial when map making is at stake. This is an important point: as discussed in Sect. 2, LPE is semiparametric and as such less prone to suffer from a misspecified model.

The heuristic motivation behind this result can be easily understood if we think of GLS map making as equivalent to OLS performed on a pre-filtered series. Pre-filtering needs not to be based on the exact form of the noise spectral density. What really matters is to strongly suppress the long range correlation and this can be achieved even with an imperfect estimate of the noise memory parameter.

3.3. Comparison with iterative methods

The purpose of this section is to compare the efficiency of LPE and WE against iterative noise-signal estimators. It is generally claimed (Prunet et al. 2001; Doré et al. 2001) that, when reasonable criteria are chosen, convergence on
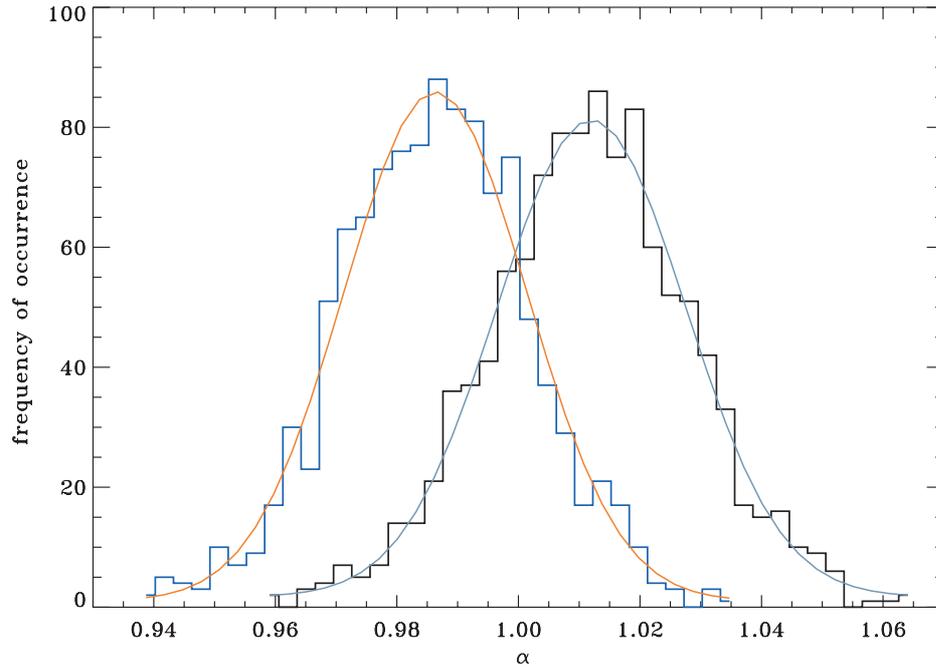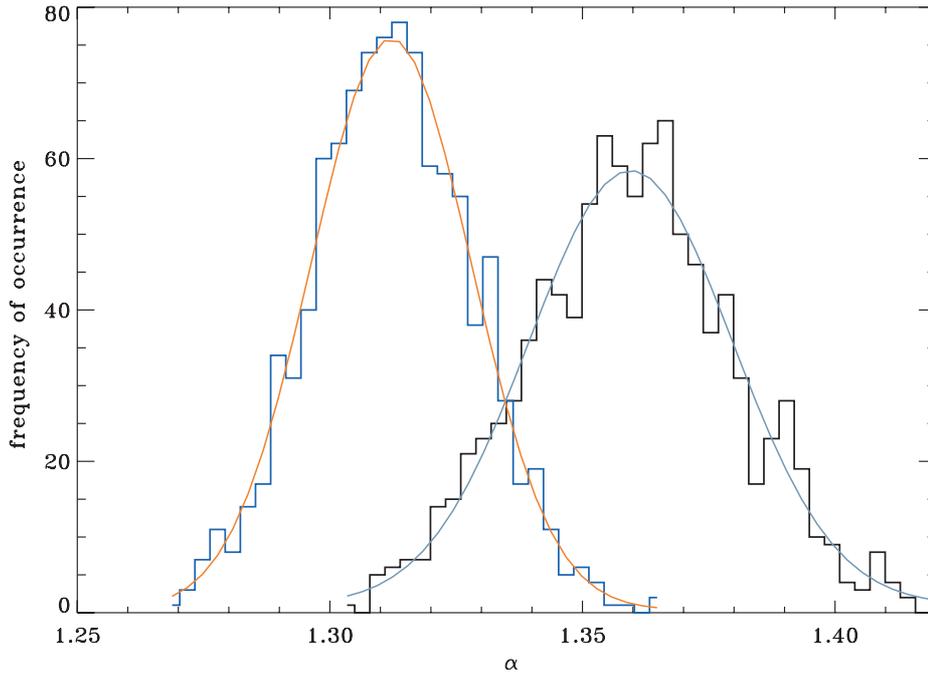
Fig. 3. Same as Fig. 1, for $\alpha = 1$.
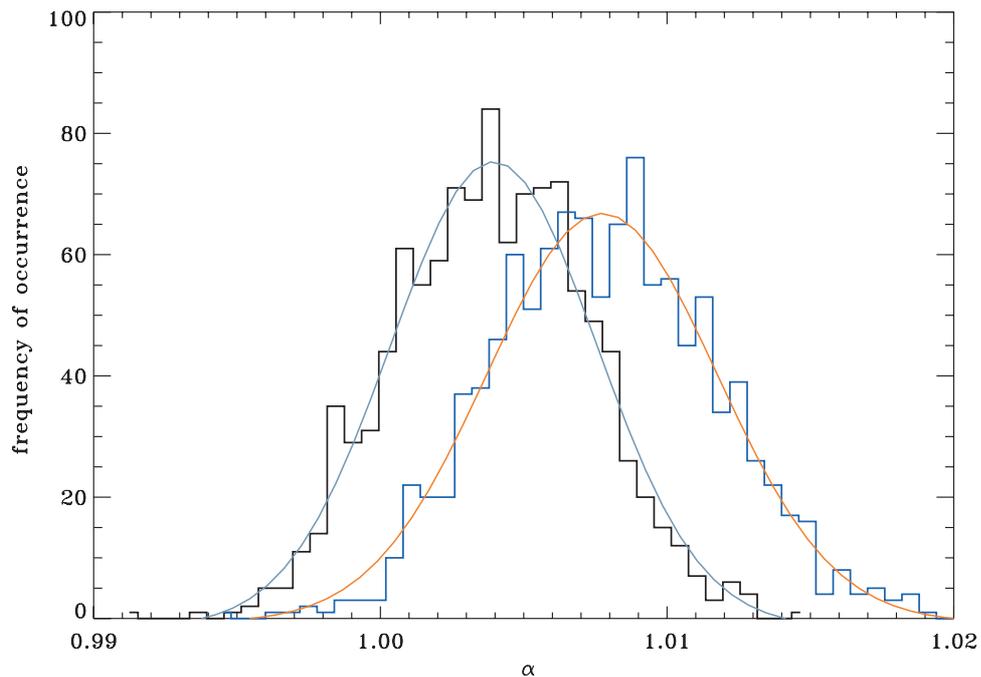
Fig. 4. Same as Fig. 2, for $\alpha = 1$.

the noise estimator is reached fairly quickly, and only a few ($\sim$4 to 5) iterations are sufficient for most applications.

The following scheme is employed to implement the iterative method. We start from the OLS noise estimator, $\tilde{\boldsymbol{n}}$. The iteration is carried out by: (1) using the latter to estimate $\boldsymbol{N}$; (2) performing map making to obtain a new solution for the map, $\hat{\boldsymbol{m}}$; (3) estimating the noise again as $\hat{\boldsymbol{n}} \equiv \boldsymbol{d} - \boldsymbol{P}\hat{\boldsymbol{m}}$. This procedure, basically the same described in Prunet et al. (2001) and more recently in Doré et al. (2001), has been implemented by using our map making

tool (Natoli et al. 2001), performing 6 iterations (1 OLS plus 5 GLS map making runs) for each map, out of a total of 50 Monte Carlo realizations. In the same spirit of Fig. 7 we derive C_ℓ's for these maps and compare them with the results based on WE (this time for an input value $\alpha = 0.7$). Results are shown in Fig. 8: the similarity between the two curves is striking.

To further compare iterative and non iterative methods we have computed histograms showing the frequency of occurrence of temperature values for selected pixels of

Fig. 5. Same as Fig. 1, for $\alpha = 1.3$.
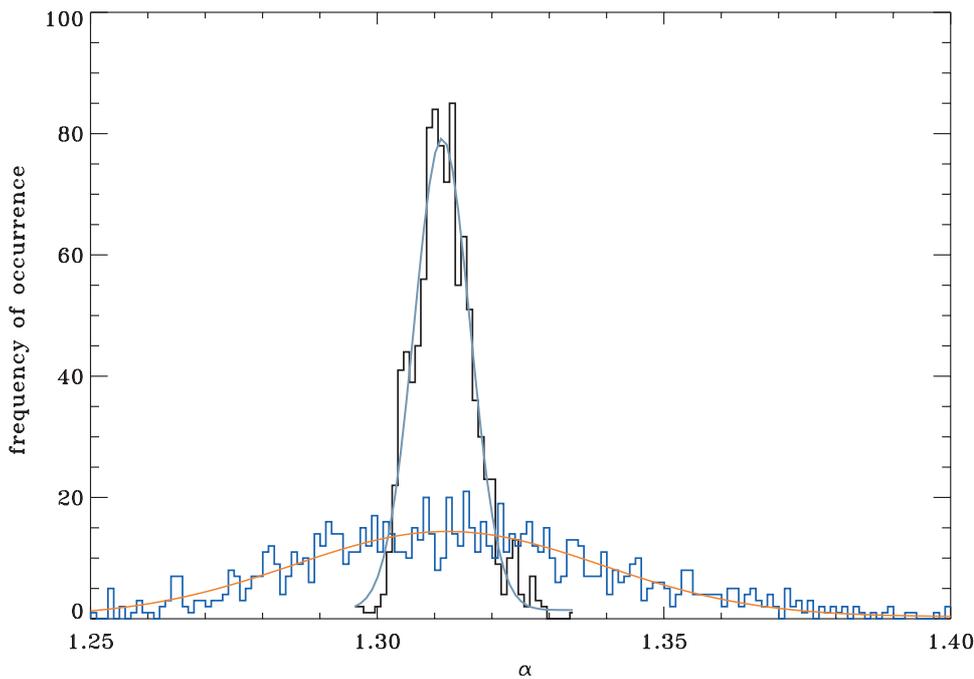

Fig. 6. Same as Fig. 2, for $\alpha = 1.3$.

the maps. We have chosen a "medium" observed pixel, which is hit 660 times in the PLANCK scan described above and a "low" observed pixel, hit 240 times. The histograms are shown in Figs. 9 and 10, respectively. Note that all the pixels with the given hit number contribute to the histograms which do not show outliers.

It should be clear at this point that estimating iteratively the noise properties yields no obvious advantage when compared to the simpler WE and LPE described

in this paper. Furthermore, as stressed in Sect. 1, the iterative scheme described above needs a sound statistical foundation in the presence of long range dependence. Moreover, it is important to realize that such a scheme is severely more costly than LPE and WE. In fact, in order to implement the latter, only a one step OLS map making is needed, plus one periodogram estimation over the time line and the final GLS one shot map making (the actual computational effort to derive WE and LPE from the
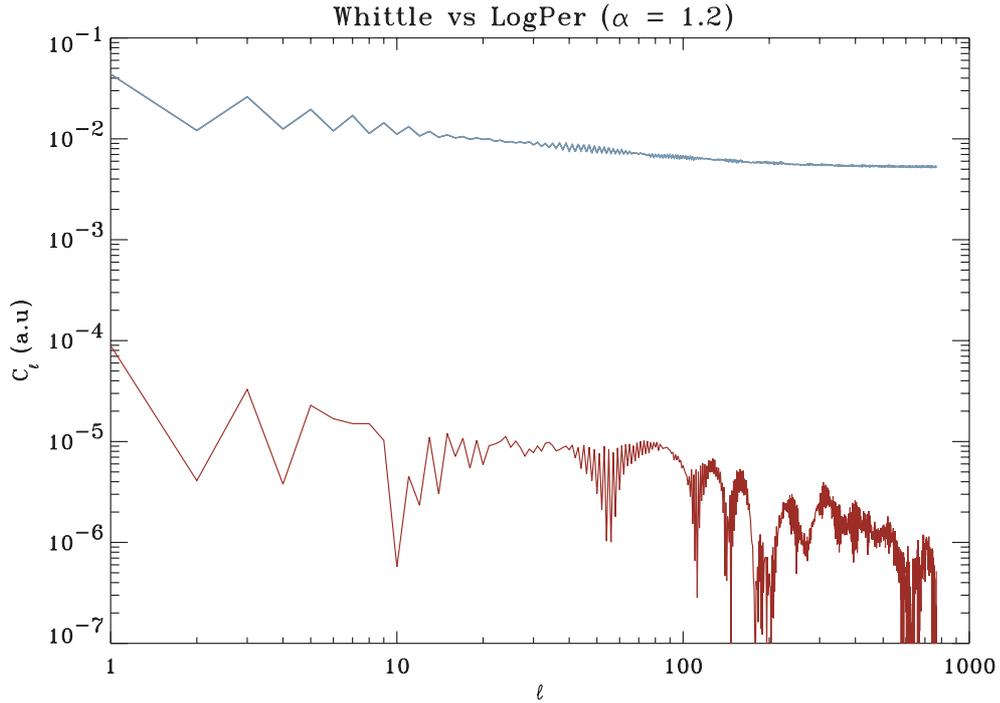
Fig. 7. Averaged angular power spectrum of 50 maps displayed as a function of multipole ℓ. The light blue line is for WE and completely hides the curve for LPE. The bottom red line is the difference between the two. The spectral index input value is $\alpha = 1.2$.
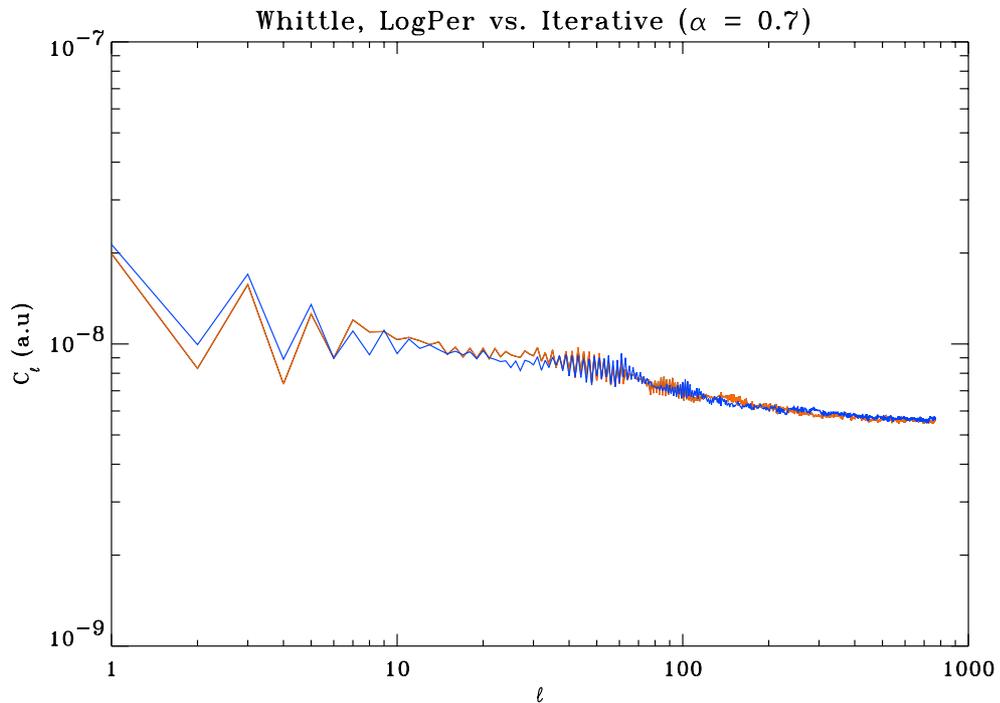

Fig. 8. As for Fig. 7 but comparing WE (blue curve) with the iterative (light red) method. The input value for α is 0.7.

periodogram is extremely small). On the other hand, the iterative method requires, other than the OLS map making, a periodogram estimation and a GLS map making at every noise iteration.

3.4. Application to BOOMERanG

All simulations shown above were performed for a subset of the PLANCK timestream. Furthermore, the noise model assumed as input is the customary "$1/f$" noise (see
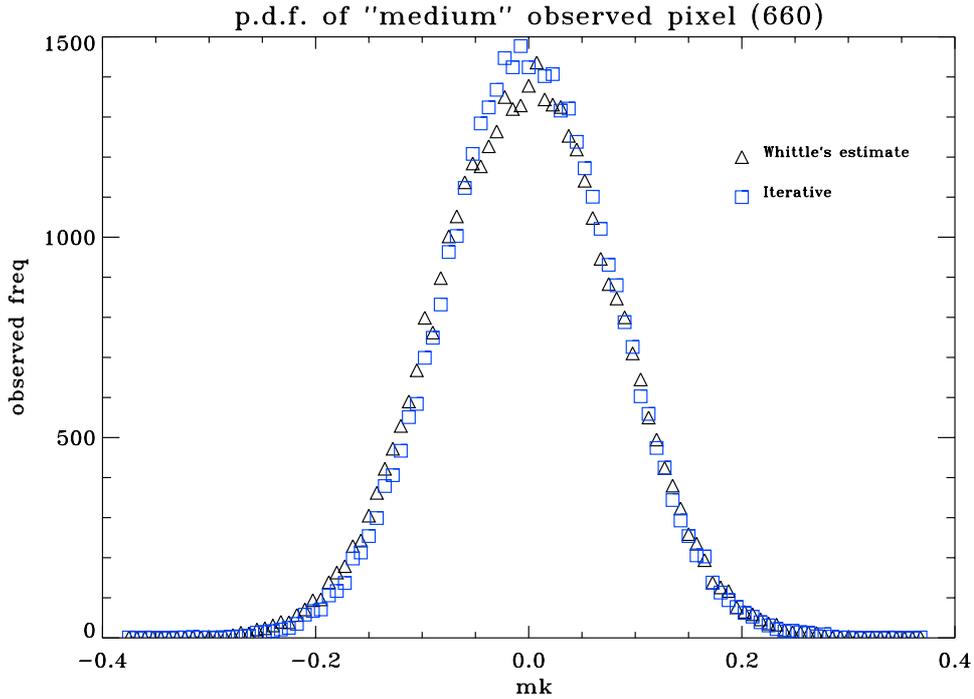
Fig. 9. Empirical distribution of the "medium" hit pixels (660 hits). WE (triangles) and the iterative scheme (squares).
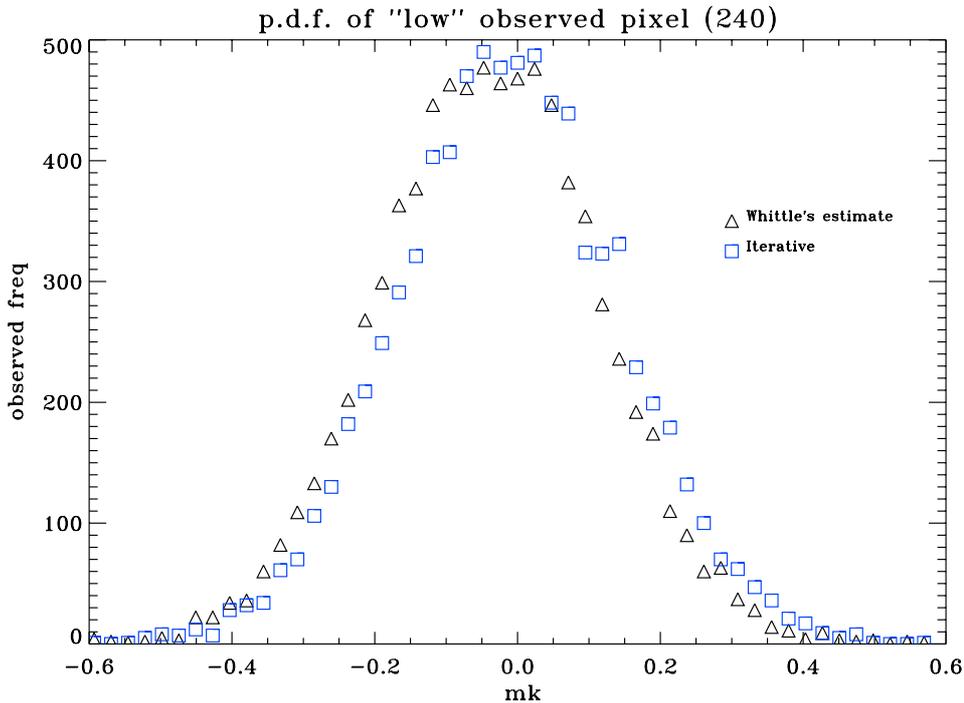

Fig. 10. Same as Fig. 9 but for pixels observed 240 times.

Eq. (6)). In this section we relax both assumptions and apply our techniques to BOOMERanG.

BOOMERanG is a balloon borne experiment which scans at constant elevation by allowing the gondola to slew at 1 or 2 degrees per second (d.p.s.), while the detector sampling rate is 60 Hz (de Bernardis et al. 1999). To make our simulations realistic we take a chunk (2^{22} samples out of total of about 3×10^7) of the 1 d.p.s. part of the scan performed by the B150A channel during the flight of 1998. Contrariously to PLANCK, this scan is aimed at a small patch. The central region comprises a few tens of square degrees and its coverage (i.e. number of hits per pixel) is far more uniform. The noise properties of the BOOMERanG detectors are not well described by the
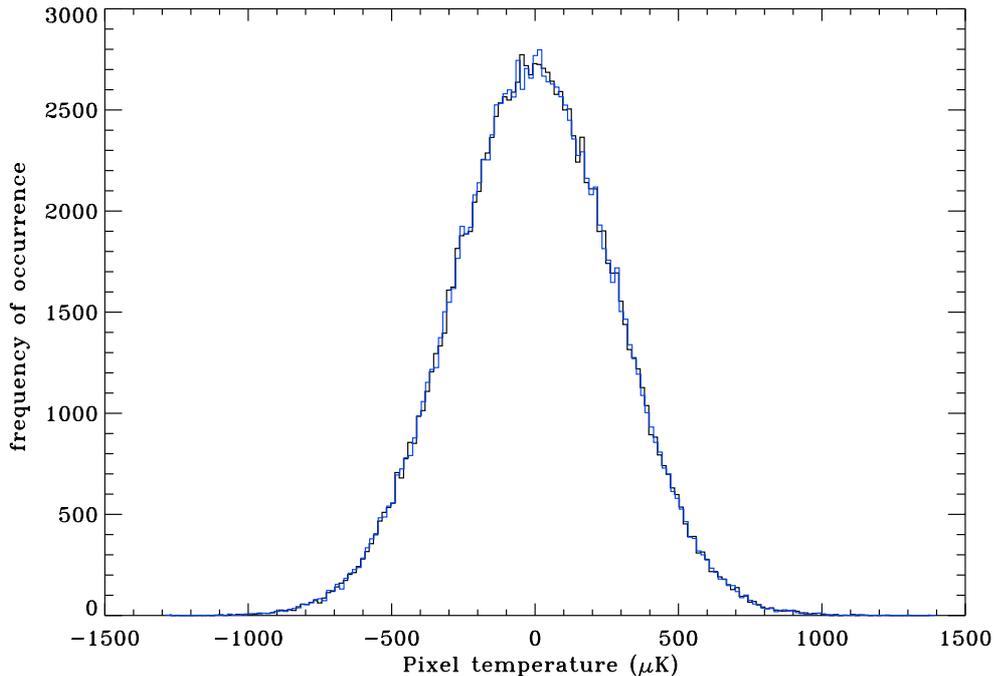
Fig. 11. Same as Figs. 9 and 10, but for a BOOMERanG-like simulation. The histogram are "observed" distributions of GLS mapmaking temperatures for all pixel hit 23 times in the maps. Noise has been estimated using WE (black curve) and LPE (dark blue).

model given in Eq. (6) because the timestream is altered to deconvolve bolometric filters (both low pass and high pass). To take these effects into account we propose to use:

$$P(f) = A \sin^2\left(\frac{\pi f}{2 f_0}\right)\left[1 + \left(\frac{f_k}{f}\right)^\alpha + A' e^{f/f_1}\right] \qquad (17)$$

with fiducial values $A \simeq 2 \times 10^4 \, (\mu K \sqrt{s})^2$, $f_0 \simeq 10^{-3}$ Hz, $f_1 \simeq 7$ Hz, $f_k = 0.15$ Hz, $A' \sim 10^8$ and $\alpha = 1.5$. We choose $f_0 \ll f_1$, since the exponential term arises after deconvolving the detector high frequency response. To avoid excessive contamination from this high frequency term we sharply cut the spectrum at 20 Hz (as in BOOMERanG: see Hivon et al. 2001).

We thus use Eq. (5) to generate Gaussian noise assuming the power spectrum given in Eq. (17). Subsequently, WE and LPE are applied to the data. WE is fully parametric and requires that a model for the spectral density is specified. Rather than supply the model given by Eq. (17), used to generate the data and hence "exact", we base the estimations on Eq. (6), i.e. on standard "$1/f''$" noise. This results in a deliberate model misspecification, as we are providing WE with an incorrect model. This choice has the obvious purpose to test the robustness of the estimate. Strictly speaking, for $\alpha \leq 2$ Eq. (17) does no longer display long memory behavior, and in this sense we are misleading LPE as well. However, long memory is still a good approximation if $f_0 \ll f_k$.

Our results are shown in Fig. 11 where we give, for the most frequently observed pixel in our scan, temperature histograms analogous to the ones given in Figs. 9 and 10. The similarity between the two curves is striking.

Very much the same conclusions as in the previous section can be drawn: the difference between WE and LPE, and between WE and the iterative method, is by all means negligible as shown in Fig. 12. In the same figure we plot the exact (unfeasible) GLS estimate. Again, no significant difference emerges.

4. Summary and conclusions

This paper has addressed estimation methods for the properties of CMB receiver noise which is well approximated by a long memory process. The most recent literature proves that optimal statistical inference is deeply affected by such characteristics. Under these circumstances results by Dahlhaus (1995) entail that map making procedures can be performed with a single iteration, provided that the slope at the origin of the log spectral density (i.e. α) is known; such procedures are asymptotically equivalent to maximum likelihood, GLS-type estimates. Our prime effort, hence, has been to address the efficient estimation of the memory parameter α. We have first presented the fully parametric WE which allows for the full correlation structure of the noise process to be recovered. WE, however, requires that the functional form of the noise spectral density is known. To relax this very restrictive assumption we also focus on the semiparametric case, where it is only assumed that a spectral singularity be present at the origin. Both the parametric and the semiparametric procedures are known to enjoy some nice robustness properties in non-standard situations, including non-Gaussian and non-stationary circumstances.
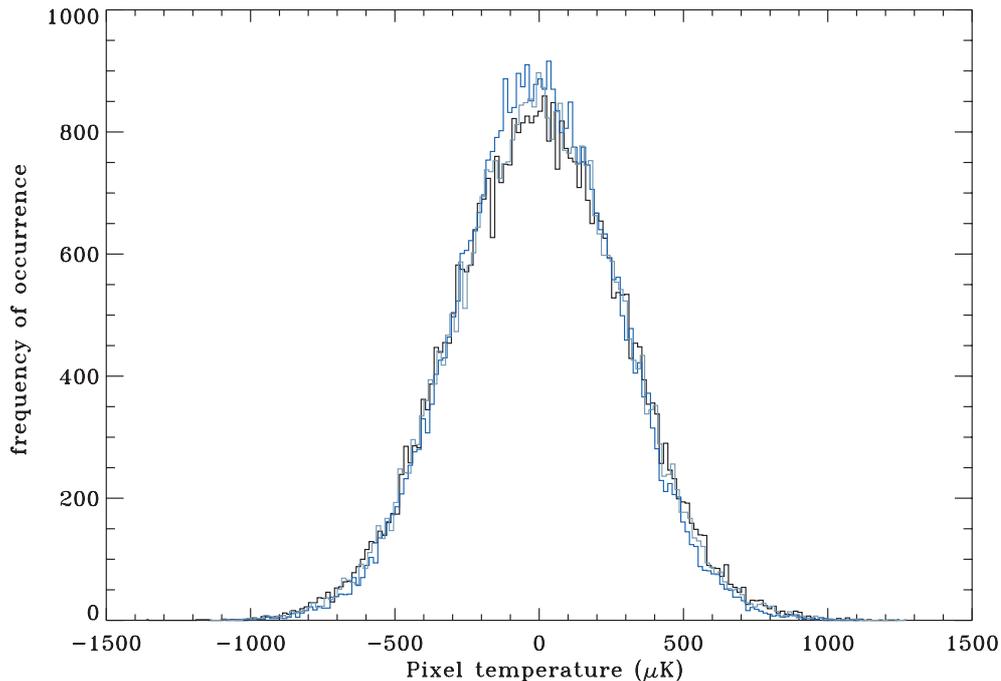
Fig. 12. Same as Fig. 11 but for a pixel hit 100 times. We compare here WE (dark blue curve) with the iterative method (black). The light blue curve is computed using the correct noise power spectrum (Eq. (17)) and, hence, provides the exact GLS estimate.

We assess the performance of these statistical procedures by implementing them on simulated PLANCK and BOOMERanG datasets. We note first that for practical implementation we need to filter out the signal component by running a preliminary OLS regression. We have shown by direct Monte Carlo comparison that this induces no significant degradation of the expected results. Our next conclusion is that the performance of semiparametric LPE is very close to correctly specified, fully parametric WE. This is comforting, as it allows for robust statistical inference under minimal assumptions. In particular, in terms of map making the improvement obtained by a full specification of the noise structure is by all means negligible. Furthermore, our simple map making procedure based on Dahlhaus (1995) ideas provides results which are virtually indistinguishable from those produced by the (computationally much more costly) iterative scheme of Prunet et al. (2001). These results are robust to implementation in the stationary and nonstationary regions.

We believe this paper leaves several avenues for further research. For example, distribution (goodness of fit) tests are profoundly affected by long memory behaviour (Dehling & Taqqu 1989; Giraitis & Surgailis 1999). Efficient estimation of α can therefore be exploited to investigate in some detail non-Gaussian features. Also, a correct evaluation of the dependence structure of receiver noise may in principle improve the determination of confidence intervals for the angular power spectrum and its governing parameters, either by resampling methods or by asymptotic approximations. These issues are currently being investigated.

References

Balbi, A., Ade, P., Bock, J., et al. 2000, ApJ, 545, L1
Bond, J. R., Jaffe, A. H., & Knox, L. 1998, Phys. Rev. D, 57, 2117
Borrill, J. 1999, Proceedings of the Conference 3 K Cosmology, AIP Conf. Proc., 476, 277
Dahlhaus, R. 1989, Ann. Statist., 17, 1749
Dahlhaus, R. 1995, Ann. Statist., 23, 1029
de Bernardis, P., et al. 1999, New Astr. Rev., 43, 289
de Bernardis, P., et al. 2001, ApJ, in press [`astro-ph/0105296`]
Dehling, H., & Taqqu, M. S. 1989, Ann. Statist., 17, 1767
Doré, O., Teyssier, R., Bouchet, F. R., Vibert, D., & Prunet, S. 2001, A&A, 374, 358
Ferreira, P. G. F., & Jaffe, A. H. 2000, MNRAS, 312, 89
Fox, R., & Taqqu, M. 1986, Ann. Statist., 14, 517
Frigo, M., & Johnson, S. G. 1998, ICASSP Conf., 3, 1381; also see `http://www.fftw.org/`
Geweke, J., & Porter-Hudak, S. 1983, J. Time Ser. Anal., 4, 221
Giraitis, L., & Surgailis, D. 1990, Probab. Theory Relat. Fields, 86, 87
Giraitis, L., & Surgailis, D. 1999, J. Statist. Plann. Inference, 80, 81

---

[5] current website: `http://www.eso.org/science/healpix`

Giraitis, L., & Taqqu, M. 1999, Ann. Statist., 27, 178

Górski, K. M., Hivon, E., & Wandelt, B. D. 1999 in Evolution of large scale structure: from recombination to Garching, ed. by A. J. Banday, R. K. Sheth, & L. N. da Costa, proc. of the MPA-ESO Cosmology conference, Garching, Germany, 2-7 August 1998, 37, PrintPartners IPSKAMP NL [astro-ph/9812350]

Hivon, E., et al. 2001, ApJ, in press [astro-ph/0105302]

Moulines, E., & Soulier, P. 1999, Ann. Statist., 27, 1415

Natoli, P., de Gasperis, G., Gheller, C., & Vittorio, N. 2001, A&A, 372, 346

Oh, S. P., Spergel, D., & Hinshaw, G. 1999, ApJ, 510, 551

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1992, Numerical Recipes in FORTRAN, The Art of Scientific Computing, 2nd Edition (Cambridge University Press, Cambridge)

Prunet, et al. 2001, to appear in proc. of the MPA/ESO conference Mining the Sky [astro-ph/0101073]

Robinson, P. M. 1995a, Ann. Statist., 23, 1048

Robinson, P. M. 1995b, Ann. Statist., 23, 1630

Robinson, P. M., & Marinucci, D. 2001, Ann. Statist., 29, 4

Tegmark, M. 1997, Phys. Rev. D, 55, 5895

Velasco, C. 1999, J. Econom., 91, 325

Velasco, C., & Robinson, P. M. 2000, J. Amer. Statist. Assoc., 452, 1229

Whittle, P. 1953, Ark. Mat., 2, 423

Wright, E. L. 1996, paper presented at the IAS CMB Data Analysis Workshop [astro-ph/9612006]