Evolution of the CMB Power Spectrum Across WMAP Data Releases: A Nonparametric Analysis

Amir Aghamousa*
Centre for Modeling and Simulation, University of Pune, Ganesh Khind, Pune 411 007 India

Mihir Arjunwadkar†
Centre for Modeling and Simulation, University of Pune, Ganesh Khind, Pune 411 007 India, and National Centre for Radio Astrophysics, University of Pune Campus, Ganesh Khind, Pune 411 007 India

Tarun Souradeep‡
Inter-University Centre for Astronomy and Astrophysics, University of Pune Campus, Ganesh Khind, Pune 411 007 India

(Dated: July 5, 2011)

In this paper, we present a comparative analysis of the Wilkinson Microwave Anisotropy Probe (WMAP) 1-, 3-, 5-, and 7-year data releases for the cosmic microwave background (CMB) angular power spectrum, with respect to the following three key questions: (a) How well is the angular power spectrum determined by the data alone? (b) How well is the ΛCDM model supported by a model-independent, nonparametric, data-driven analysis? (c) What are the realistic uncertainties on peak/dip locations and heights? Our analysis is based on a model-independent nonparametric function estimation methodology [1, 2]. A unique feature of this methodology is that it provides, in addition to the fit, a powerful inferential object, viz., a confidence set around the fit. Questions of interest to cosmology, suitably translated in terms of features of the fit, can be meaningfully addressed using the confidence set. For example, this construct allows for simultaneous probing of any number of features (such as locations and heights of peaks and dips) of the fit. It can also be used to test the credibility of fits obtained using other statistical methods. Our results show that the height of the power spectrum is well determined by data alone for $l \lesssim l_d$, with approximate upper bounds on l_d being 600 (1-year), 800 (3-year), and 900 (5- and 7-year data realizations). We also show that parametric fits based on the ΛCDM model are remarkably close to our nonparametric fit in l-regions where the data are sufficiently precise. A contrasting example is provided by an HΛCDM model: As the data become more and more precise with successive data realizations, the HΛCDM angular power spectrum gets progressively pushed away from our nonparametric fit, indicating the incompatibility of this particular cosmological model with respect to the WMAP data sets. We present uncertainties on peak/dip locations and heights at the 95% (2σ) level of confidence, and show how these uncertainties translate into hyperbolic "bands" on the acoustic scale (l_A) and the peak shift (ϕ_m) parameters. Based on the confidence set for the fit to the 7-year data, we also argue that the low-l up-turn in the CMB angular power spectrum cannot be ruled out at any confidence level in excess of about 10% ($\approx 0.12\sigma$). Additional outcomes of this work are a numerical formulation for risk minimization subject to monotonicity constraints for a noise-weighted risk function and a prescription for obtaining nonparametric fits that are closer to cosmological expectations on smoothness, which extend the utility of this nonparametric methodology in the context of cosmological function estimation problems.

PACS numbers: 98.70.Vc, 98.80.Es, 98.80.Cq, 95.75.Pq, 02.50.Tt
Keywords: Cosmic microwave background (CMB), angular power spectrum, Wilkinson microwave anisotropy probe (WMAP) data, cosmological parameter estimation, nonparametric regression and inference, statistical data analysis

I. INTRODUCTION

The angular power spectrum of cosmic microwave background (CMB) temperature fluctuations is a measurable physical quantity that depends sensitively on the physics of the early universe. In particular, the shape of the angular

*Electronic address: amir@cms.unipune.ac.in
†Electronic address: mihir@cms.unipune.ac.in
‡Electronic address: tarun@iucaa.ernet.in

power spectrum and the locations and heights of its peaks relate directly to parameters in the underlying cosmological models. As such, it has been used extensively as an acid test of the relative merit of competing cosmological models, and as a rich information source on cosmological parameters themselves.

Traditionally, and almost exclusively, cosmologists have resorted to model-based parametric statistical methods for estimating the CMB angular power spectrum from data. Parametric regression methods require the functional form of the unknown regression function f to be pre-specified. The values of adjustable parameters in f, finite in number, are usually estimated by maximizing an appropriate likelihood function or a posterior distribution. In the cosmological context, it is conventional to employ parametric models that attempt to capture the essential physics of the problem via the pre-specified functional form, and any pre-existing knowledge about parameters is incorporated in the estimation process via appropriate prior distributions.

Nonparametric function estimation methods, on the other hand, assume no specific functional form for f (except for mild regularity conditions such as smoothness assumptions, membership to a function space, etc). In this approach, an estimate $\widehat{f}_N$ of the unknown regression function f is obtained by balancing bias and variance of $\widehat{f}_N$ via optimal smoothing. Nonparametric methods are therefore model-independent, and are based on fewer and less restrictive assumptions about f. This, in turn, implies that any inferences about f made from nonparametric regression analyses tend to be more data-driven (in contrast to being model-driven). In other words, to a greater extent nonparametric function estimation methods tend to infer what *is* rather than what *should be*. As such, nonparametric regression methods can be meaningfully employed as sanity-enforcing mechanisms on parametric analyses, thereby making the conclusions drawn more conservative. For example, a feature seen in a parametric fit that survives in a nonparametric analysis is likely to be a real and robust feature of the data itself, and not merely an artifact that is seen because a parametric model expects it to be there. Alternative methodologies, such as the nonparametric methodology [1, 2] used in this work, may allow posing inferential questions that are difficult to address using conventional methods. Specifically, this methodology has the formal advantage of being able to provide realistic uncertainties on *any* number of features of the angular power spectrum that are simultaneously valid at the same level of confidence. Such desirable features are arguably lacking [1] in most mainstream methodologies, including Bayesian, that are commonly used in cosmology.

The four CMB angular power spectrum data sets [3–6] released by the Wilkinson Microwave Anisotropy Probe (WMAP) [7] mission, representing cumulative observations at the end of 1, 3, 5, and 7 years of operation, present a unique opportunity for statistical analysis. For example, till date these are claimed to represent the most precise and extensive full-sky CMB measurements ever made (only to be superseded by the Planck mission [8]). The four data sets represent the evolution of data over a period of about a decade, thereby making it possible to assess progressive and possibly systematic resolution of features of the spectrum. From a statistical perspective, each of these moderately large data sets (minimum is 899 data points for WMAP 1-year) is not only heteroskedastic, but also has substantial correlations that arise in typical data pipelines [9–13].

In this paper, we present a comparative nonparametric analysis of the WMAP 1-, 3-, 5-, and 7-year data sets for the angular power spectrum. Specifically, for each data realization, we address the following three key questions: (a) How well is the angular power spectrum determined by the data alone? (b) How well is the ΛCDM model supported by a model-independent, nonparametric, data-driven analysis? (c) What are the realistic uncertainties on peak/dip locations and heights? Our analysis is based on a nonparametric function estimation methodology [1, 2], which is discussed in Sec. II together with our extensions for monotone minimization of the inverse-noise-weighted risk function, and for obtaining a monotone nonparametric fit that is closer to cosmological expectations for smoothness. Results are presented in Sec. III, and conclusion in Sec. IV.

II. METHODOLOGY

The nonparametric function estimation methodology [1, 2] used in this work is an extension of the REACT methodology [14, 15], which is in turn founded on rigorous formal results in [16, 17]. A somewhat elementary treatment of this method can be found in [18]. Two early papers [19, 20] used this methodology to analyze, under the assumption of homoskedasticity, a pre-WMAP data set that combined BOOMERanG, MAXIMA, and DASI data sets. A generalization of this formalism for dealing with heteroskedasticity via an inverse-noise-weighted loss function was developed in [1]. More recently, using the WMAP 1-year data, [2] illustrated how confidence intervals on cosmological parameters and boundaries in the cosmological parameter space can be inferred from the confidence set for a nonparametric power spectrum fit. An incisive and discerning discussion about the relative merits of this methodology over conventional methods can be found in [1].

In this section, we present an operational outline of this methodology (Sec. II A and II B). This outline is based on [1, 2] and is included for completeness. Our own numerical formulation of the monotone risk minimization problem, where the risk function is derived from an inverse-noise-weighted loss function, is presented in Sec. II C. In Sec. II D,

we describe a systematic way of obtaining a monotone fit with smoothness that meets cosmological expectations.

A. The nonparametric fit

We are given CMB angular power spectrum data of the form

$$Y_l = C_l + \epsilon_l \tag{1}$$

consisting of N data points observed over multipole index range $l_{min} \leq l \leq l_{max}$. Here, C_l stands for the value of the true but unknown power spectrum at multipole index l. The noise variables ϵ_l are assumed to have a mean-0 normal distribution with *known* covariance matrix Σ. In practice, any reasonable estimate/approximation $\widehat{\Sigma}$ of this covariance matrix, such as an inverse Fisher matrix for a pilot parametric fit, can be used in place of Σ [1].

This nonparametric regression method is based on expanding the unknown regression function f, assumed to belong to an appropriate L_2 function space, in a complete orthonormal basis $\{\phi_j(x)\}$, as

$$f(x) = \sum_{j=0}^{\infty} \beta_j \phi_j(x).$$

A basis that has proven useful in the CMB angular power spectrum context is the cosine basis defined over $0 \leq x \leq 1$:

$$\phi_j(x) = \begin{cases} 1 & (j = 0) \\ \sqrt{2}\cos(j\pi x) & (j = 1, 2, \ldots) \end{cases} . \tag{2}$$

Assuming that f is sufficiently smooth, we take

$$f(x) \approx \sum_{j=0}^{N-1} \beta_j \phi_j(x),$$

and estimate it as

$$\widehat{f}_N(x) = \sum_{j=0}^{N-1} \widehat{\beta}_j \phi_j(x). \tag{3}$$

While the method is asymptotically basis-independent, choice of the basis may matter in any finite-data-size application; see [14, 15] for a detailed discussion. This basis satisfies a discrete orthogonality property when the data Y_i are available over an equispaced grid $\{x_i = (2i+1)/2N, 0 \leq i \leq N-1\}$ consisting of zeros of $\phi_N(x)$. In the CMB context, any contiguous range of N integer multipole indices $l_{min} \leq l \leq l_{max}$ can be formally mapped onto this equispaced grid, hence we will not make any categorical distinction between data index i and the corresponding multipole index l.

The true angular power spectrum $C_l \equiv f(x_l)$ is estimated as $\widehat{C}_l \equiv \widehat{f}_N(x_l)$ via coefficient estimates $\widehat{\beta}_j$, which are estimated as

$$\widehat{\beta}_j = \lambda_j Z_j, \tag{4}$$

where

$$Z_j = \frac{1}{N} \sum_{i=0}^{N-1} Y_i \phi_j(x_i) = \frac{(U^T Y)_j}{\sqrt{N}}. \tag{5}$$

Here, U is the orthonormal matrix with elements $U_{ij} = \phi_j(x_i)/\sqrt{N}$ and $Y \equiv (Y_0, \ldots, Y_{N-1})^T$.

The task of obtaining coefficient estimates $\widehat{\beta} \equiv (\widehat{\beta}_0, \ldots, \widehat{\beta}_{N-1})^T$, and thereby the fit $\widehat{f}_N(x_i)$, is now relegated to determining the *shrinkage parameters* λ_j. Assuming smoothness for f (which implies a rapid decay of the true coefficients β_j with j), the shrinkage parameters λ_j are constrained to be monotonically decreasing

$$1 \geq \lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{N-1} \geq 0 \ \text{(Monotone shrinkage)}. \tag{6}$$

A special, discrete subset of the monotone shrinkage defined above is the *nested subset selection* (NSS) shrinkage [1], defined as

$$\lambda_j = \begin{cases} 1 \text{ for } 0 \le j < J \\ 0 \text{ for } J \le j < N \end{cases} \quad \text{(NSS shrinkage)}. \tag{7}$$

Either shrinkage type results in selective damping of high-frequency harmonics in the data Y_i, which results in smoothing of the fit $\widehat{f}_N$. A useful interpretation of shrinkage parameter λ_j is that it represents the *effective degree of freedom* for the jth coefficient estimate $\widehat{\beta}_j$. One can thus define the effective degrees of freedom (EDoF) for the entire fit $\widehat{f}_N$ as [37]

$$\text{EDoF}(\lambda) = \sum_{j=0}^{N-1} \lambda_j. \tag{8}$$

Optimal smoothing is obtained by balancing the bias of the fit $\widehat{f}_N$ with its variance, by minimizing the risk function

$$\widehat{R}(\lambda) = Z^T \bar{D} W \bar{D} Z + \text{tr}(DWDB) - \text{tr}(\bar{D} W \bar{D} B) \tag{9}$$

with respect to the shrinkage parameters. Monotone shrinkage usually results in a lower risk than the NSS shrinkage because of the greater freedom available in the monotone set of shrinkages. This risk function is an estimator for the expected value of the inverse-noise-weighted loss function

$$L(\lambda) = \int_0^1 \left(\frac{f(x) - \widehat{f}_N(x)}{\sigma(x)} \right)^2 dx.$$

Here, $\lambda \equiv (\lambda_0, \ldots, \lambda_{N-1})^T$ is the vector of shrinkage parameters, $\sigma^2(x)$ is the (known) variance of the data Y at x, $Z \equiv (Z_0, \ldots, Z_{N-1})^T$, $D \equiv \text{diag}(\lambda_0, \ldots, \lambda_{N-1})$, $\bar{D} = I_N - D$, $B = U^T \Sigma U / N$ is the covariance of Z, and I_N is the $N \times N$ identity matrix. The positive (semi)definite weight matrix W is defined as

$$W_{jk} = \sum_l \Delta_{jkl} w_l, \tag{10}$$

where w_l is the lth coefficient in the expansion $(1/\sigma^2(x)) \approx \sum_{j=0}^{N-1} w_j \phi_j(x)$ and, for the cosine basis (Eq. 2),

$$\Delta_{jkl} = \int_0^1 \phi_j(x)\phi_k(x)\phi_l(x)\, dx = \begin{cases} 1, & \text{if } \#\{j,k,l=0\} = 3, \\ 0, & \text{if } \#\{j,k,l=0\} = 2, \\ \delta_{jk}\delta_{0l} + \delta_{jl}\delta_{0k} + \delta_{kl}\delta_{0j}, & \text{if } \#\{j,k,l=0\} = 1, \\ \frac{1}{\sqrt{2}}(\delta_{l,j+k} + \delta_{l,|j-k|}) & \text{if } \#\{j,k,l=0\} = 0. \end{cases}$$

We denote the optimal shrinkage obtained by minimizing risk $\widehat{R}(\lambda)$ by $\widehat{\lambda}$. The best NSS shrinkage $\widehat{\lambda}_{NSS}$ is obtained simply by evaluating risk $\widehat{R}(\lambda)$ for each of the $(N+1)$ NSS shrinkage vectors and choosing the one with the least risk. We will discuss risk minimization subject to monotonicity constraints (Eq. 6) in Sec. II C.

Fig. 1 illustrates the contrasting behavior of the nonparametric risk (red curve) and the WMAP 1-year likelihood function [21] (green curve) for the WMAP 1-year data [3], as a function of the EDoF of all NSS fits. This figure is motivated by the fact that cosmologists, by and large, are better-acquainted with parametric likelihood-based methods. The vertical axis represents the two quantities shifted and scaled to a common range of 0 to 100. Each integer value on the horizontal axis represents one NSS fit, from the zero function at EDoF = 0 to the fit that simply interpolates through the data (EDoF = N). Optimal smoothing for NSS shrinkage occurs at EDoF = 12 where the nonparametric risk function attains its minimum over the NSS set of fits. Likelihood function, on the other hand, keeps on improving with the EDoF indefinitely.

B. Confidence set around the fit

Conventional regression methods provide a *confidence band* around the fit that tries to quantify the uncertainty in the fit. In contrast, this nonparametric methodology quantifies the uncertainty surrounding the nonparametric fit in

the form of an elegant construct, namely, a $(1-\alpha)$ *confidence set* at a pre-specified confidence level $0 \leq (1-\alpha) \leq 1$. The $(1-\alpha)$ confidence set for the coefficient vector β is defined as

$$\mathcal{D}_{N,\alpha} = \left\{ \beta : (\beta - \widehat{\beta})^T W (\beta - \widehat{\beta}) \leq r_\alpha^2 \right\}, \tag{11}$$

which is centered at the vector of estimated coefficients $\widehat{\beta}$, and the *confidence radius* r_α is given by

$$r_\alpha^2 = \frac{\widehat{\tau} z_\alpha}{\sqrt{N}} + \widehat{R}(\widehat{\lambda}). \tag{12}$$

Here, z_α is the upper α quantile of the standard normal distribution, and

$$\widehat{\tau}^2/N = 2\mathrm{tr}(ABAB) + Z^T Q Z - \mathrm{tr}(QB), \tag{13}$$

with $Q = 4(ABA + WDBDW - 2ABDW)$ and $A = DW + WD - W$ [38]. The corresponding confidence set on the true regression function f is given by

$$\mathcal{B}_{N,\alpha} = \left\{ f(x) = \sum_{j=0}^{N-1} \beta_j \phi_j(x) : \beta \in \mathcal{D}_{N,\alpha} \right\}. \tag{14}$$

The quadratic form of the inverse noise-weighted loss function and the fact that the weight matrix W is positive (semi)definite implies that both confidence sets $\mathcal{D}_{N,\alpha}$ and $\mathcal{B}_{N,\alpha}$ are ellipsoidal in shape. For any functional T of the spectrum f, such as location or height of a peak or a dip, the $(1-\alpha)$ confidence interval is defined as

$$\mathcal{I}_{N,\alpha} = \left(\min_{f \in \mathcal{B}_{N,\alpha}} T(f), \max_{f \in \mathcal{B}_{N,\alpha}} T(f) \right). \tag{15}$$

Moreover, prior information that the true regression function f belongs to a subset $\mathcal{P}_{N,\alpha}$ of the confidence set $\mathcal{D}_{N,\alpha}$ (e.g., f has k peaks over the range of x-values represented in the data) can be incorporated in the analysis by replacing $\mathcal{D}_{N,\alpha}$ with $\mathcal{P}_{N,\alpha} \cap \mathcal{D}_{N,\alpha}$. This methodology further provides the formal assurance that, asymptotically,

1. $\mathcal{B}_{N,\alpha}$ ($\mathcal{D}_{N,\alpha}$) will contain the true spectrum f (true coefficient vector β) with probability $\geq (1-\alpha)$, and
2. confidence intervals $\mathcal{I}_{N,\alpha}$ on *any number of* functionals $T(f)$, computed from the confidence set $\mathcal{B}_{N,\alpha}$, will be *simultaneously valid* at the same confidence level $(1-\alpha)$, and that these will trap their corresponding true but unknown values with probability $\geq (1-\alpha)$.

C. Risk minimization subject to monotonicity constraints

In this section, we show how the risk function $\widehat{R}(\lambda)$ (Eq. 9) can be minimized subject to the monotonicity constraints $1 \geq \lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{N-1} \geq 0$. The risk function corresponding to the unweighted loss function ($W = I_N$) has a simple weighted-sum-of-squares form, and can be minimized exactly and efficiently using the pooled adjacent violators (PAV) algorithm [22]. While the risk function corresponding to the inverse-noise-weighted loss function ($W \neq I_N$) is still quadratic in λ, it can no longer be expressed as a weighted sum-of-squares, and the PAV algorithm cannot be used to minimize it.

It can be shown that, disregarding terms that do not depend on λ, the risk function $\widehat{R}(\lambda)$ (Eq. 9) can be written as

$$\widehat{R}(\lambda) = \frac{1}{2} \lambda^T H \lambda - \lambda^T h,$$

where $H_{jk} = 2 z_j z_k W_{jk}, h = (H - V)(1, 1, \ldots, 1)^T$, and $V_{jk} = 2 W_{jk} B_{kj}$. H and V are both manifestly symmetric. Positive (semi)definiteness of W implies that H is a positive (semi)definite matrix, implying that $\widehat{R}(\lambda)$ is a convex function. The system of linear inequality constraints $1 \geq \lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{N-1} \geq 0$ implies that the constrained region (Fig. 2) has a convex trianguloidal shape determined by flat surfaces. The original minimization problem can therefore be formulated as follows:

$$\begin{aligned} \text{Minimize} \quad & \widehat{R}(\lambda) = \frac{1}{2} \lambda^T H \lambda - \lambda^T h \\ \text{subject to} \quad & C\lambda \leq (0, 0, \ldots, 0)^T \\ \text{and} \quad & 0 \leq \lambda_i \leq 1 \ \text{ for all } i, \end{aligned} \tag{16}$$

where C is the $(N-1)\times N$ matrix

$$C = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}.$$

An additional linear inequality or equality constraint of the form

$$\sum_{i=0}^{N-1} \lambda_i \leq q \tag{17}$$

or

$$\sum_{i=0}^{N-1} \lambda_i = q, \tag{18}$$

where q constrains the EDoF of the fit ($0 < q \leq N$), can easily be accommodated in this formulation. This reformulation of the monotone risk minimization problem makes it possible to use standard minimization methods [23–25] and computational tools [26, 27] for convex quadratic problems subject to linear constraints and simple bounds.

D. Obtaining a smoother monotone fit that is closer to cosmological expectations

To motivate the discussion in this section, consider the full-freedom monotone fit to the WMAP 7-year data (green curve in the WMAP 7 panel; Fig. 4) obtained as above. By *full-freedom monotone fit*, we mean the fit that minimizes risk over the entire monotone-admissible region (Fig. 2) without any restriction on the EDoF of the fit. This fit turns out to be quite wiggly especially at the high-l end because of the high noise variance here. Such wiggliness implies presence of high-frequency components in the fit, which in turn implies a large number of EDoF in the fit.

Without cosmological pre-conditioning (i.e., from a completely agnostic and data-driven perspective) and when viewed in relation to the data, it is clear that this fit is not at all unreasonable, especially given the high noise levels at the high-l end. However, all cosmological models suggest far smoother shapes for the angular power spectrum. In the context of the present methodology, one candidate for a smoother fit is the NSS fit (i.e., one that has the minimal risk over the set of $(N+1)$ NSS fits). Indeed, this possibility has been exploited, e.g., in [1]. However, the NSS fit (see the blue curve in Fig. 4) may also turn out to be somewhat unsatisfactory with respect to cosmological expectations (and, some times, also with respect to trends reflected in the full-freedom monotone fit). This is primarily because of the limited freedom available in the NSS class. Notice again that the NSS fit is not entirely unreasonable from an agnostic viewpoint.

The monotone set, on the other hand, offers the possibility of harnessing local minima in the risk function that are constrained to lie in appropriate "smoother" subsets of the full monotone set. This may be achieved in two distinct ways that can be combined for greater effect:

1. By imposing one of the additional constraints, Eq. 17 or 18, on the EDoF of the fit. Examples of such restricted-freedom monotone fits are the red curves in Fig. 4.

2. By truncating the expansion (Eq. 3) to p number of coefficients ($p < N$) and then performing monotone risk minimization over this subset of the full monotone set.

In practice, such smoother restricted-freedom monotone fit can be obtained by gradually reducing the value of q (Eq. 17 or 18) starting from the EDoF of the NSS fit until all low-amplitude, high-frequency wiggles in the fit disappear. Generally, the resulting fit has a lower risk than the NSS fit with EDoF $= q$, and is manifestly consistent with trends captured by the full-freedom monotone fit. We find it useful to present (or consider) all three fits (NSS, full-freedom monotone, and restricted-freedom smoother monotone) together: This helps build a realistic picture about estimated trends in the data, and thereby about the shape of the underlying true spectrum. Like the NSS fit, the smoother restricted-freedom monotone fit will generally be more biased than the full-freedom monotone fit. This greater bias, however, is partially compensated for by a larger risk which results in a larger confidence radius value (Eq. 12), a larger confidence set, and therefore more conservative inferences.

III. RESULTS AND DISCUSSION

The four WMAP angular power spectrum data sets used in this work, ΛCDM parametric fits for the CMB angular power spectrum, and likelihood codes that produce their respective Fisher (inverse covariance) matrices are obtained from the WMAP data archive `http://lambda.gsfc.nasa.gov/`. For all four data realizations, it turned out that (a) the weight matrix W (Eq. 10) is numerically positive definite, and (b) the confidence set for the full-freedom monotone fit is completely nested inside that for the smoother restricted-freedom monotone fit. It is worth noting that our nonparametric fits and confidence sets are not too sensitive to the details of the covariance matrix (this was also pointed out in [2]). Most computations were done using the `R` statistical computing environment [28]. We used the `QL` codes [26] for the monotone risk minimization problem (Eq. 16). Our nonparametric fits, obtained using the method outlined in Sec. II C, are shown in Fig. 4 for the WMAP 1-, 3-, 5-, and 7-year data sets.

A. How well is the power spectrum determined by data alone?

Considering that the noise is highly heteroskedastic and noise levels are high for large l, it is useful to make an assessment of how noise in the data affects local uncertainties in the fitted spectrum, and to quantify how well the angular power spectrum value at each l is determined by the data. This can be done approximately using the boxcar probe approach [1]. For this purpose, it is convenient to use the following approximate form of the confidence set (Eq. 14)

$$\mathcal{B}_{N,\alpha} \approx \left\{ (C_{l_{min}}, \ldots, C_{l_{max}}) : \frac{1}{N} \sum_{l=l_{min}}^{l_{max}} \left(\frac{C_l - \widehat{C}_l}{\sigma_l} \right)^2 \leq r_\alpha^2 \right\},$$

where $\widehat{C}_l$ and σ_l^2 are respectively the fitted value of the power spectrum and the noise variance at l. In this approach, we generate variations $(C_{l_{min}}, \ldots, C_{l_{max}})$ of the power spectrum by adding flat localized perturbations (boxcars) of fixed l-width to the fit. The height of the boxcar is determined by the condition that the resulting variation remains on the surface of the 95% confidence set $\mathcal{B}_{N,\alpha}$ (Eq. 14).

In Fig. 3, we plot this height, scaled by the value of the fit, as a function of the multipole index l, for all four data realizations. This is interpreted as follows: Values $\ll 1$ indicate that the fit is well determined by the data, whereas values > 1 imply that the data contain little or no information about the height of the power spectrum. We see that the range of l-values over which the fit is well-determined expands consistently between 1-, 3- and 5-year data realizations, from $l \approx 600$ to $l \approx 900$. On the other hand, while the l-range of the data expanded between WMAP 5 and 7, the information contained in the data does not appear to have grown proportionately beyond $l \approx 900$.

B. How well is the ΛCDM model supported by a model-independent, nonparametric, data-driven analysis?

In each of the four plots in Fig. 4, we have also included parametric fits based on the ΛCDM [3–6] and HΛCDM [29, 30] models (see figure caption for details). The parametric ΛCDM-based fits turn out to be quite close to the respective nonparametric fits wherever the data are precise. This is remarkable considering that our nonparametric method does not rely on any cosmologically-motivated prior information whatsoever. Moreover, the parametric ΛCDM-based fit appears to get closer to the respective nonparametric fit across the four WMAP data releases. The closeness of a parametric fit $(C_{l_{min}}, \ldots, C_{l_{max}})$ to the corresponding nonparametric fit $(\widehat{C}_{l_{min}}, \ldots, \widehat{C}_{l_{max}})$ can be measured through the distance

$$d(C, \widehat{C}) \approx \sqrt{\frac{1}{N} \sum_{l=l_{min}}^{l_{max}} \left(\frac{C_l - \widehat{C}_l}{\sigma_l} \right)^2}.$$

Using Eq. 12, this distance can be further expressed as the smallest confidence level α beyond which the parametric fit is rejected as a candidate for the true spectrum.

Table I lists distances of parametric fits based on the two cosmological models (ΛCDM and HΛCDM) from the nonparametric full-freedom monotone fit for the corresponding data realization (see caption for specific details and description). The progression of distance values between a parametric ΛCDM fit and the corresponding nonparametric fit clearly shows that the two are getting closer as the data become precise. In contrast to this, the angular power spectrum generated by the particular HΛCDM model considered, which is almost as strong a contender for the true

power spectrum as the ΛCDM fit with respect to the 1-year confidence set, is progressively pushed away to the boundary of the 95% confidence set for the 7-year confidence set. Visually, this trend can be understood on the basis of the differences between the HΛCDM power spectrum and the nonparametric fit (e.g., differences in the heights of the first peak; see Fig. 4) that result into pushing this particular HΛCDM model out of the confidence set. Given the formal guarantees of this methodology (Sec. II B), the WMAP 7-year data thus rules out, at $\approx 95\%$ confidence level, the particular HΛCDM model considered.

C. Uncertainties on locations and heights of peaks and dips

We now consider the problem of determining uncertainties on the locations and heights of peaks and dips in the nonparametrically fitted spectrum. The motivation for this exercise comes from the fact that peak locations and heights contain valuable information about cosmological models and parameters [31, 32]. Our particular method for probing the confidence set for determining uncertainties in peak/dip locations and heights is based on the following observations:

1. The confidence set $\mathcal{D}_{N,\alpha}$ on the vector of coefficients $(\beta_0, \ldots, \beta_{N-1})$, by construction, is centered at the vector of estimates $(\widehat{\beta}_0, \ldots, \widehat{\beta}_{N-1})$.

2. The confidence interval defined in Eq. 15 requires locating extreme variations in any functional T of the power spectrum f (such as peak location or height). The largest possible variations in T will be located as far away from the center of the confidence set as possible, i.e., on its surface.

3. Cosmologically meaningful and sufficiently smooth variations in the fitted spectrum $\widehat{C}_l$ are most likely to be located in the projection of the full confidence set onto the lowest M dimensions.

We therefore generate a uniform sample from the projection of the full confidence set surface onto the lowest d dimensions, where $2 \leq d \leq M$, with $M \lesssim 23$. For convenience, we use the smoother restricted-freedom monotone fit for this purpose, with the justification that the confidence set corresponding to the full-freedom monotone fit happens to be nested inside that for this fit, for all four data realizations. The Choleski factorization $W = u^T u$ is used to transform the original confidence ellipsoid $\mathcal{D}_{N,\alpha}$ (whose principal axes may not be aligned with coordinate directions in the β-space) into a sphere of the form $\{\psi : \|\psi - \widehat{\psi}\|^2 \leq r_\alpha^2\}$, where $\psi = u\beta$ and $\widehat{\psi} = u\widehat{\beta}$. The surface of this ψ-sphere can be efficiently sampled with uniform density using a standard algorithm (Sec. 3.4.1.E.6 on p.130 of [33]), and then transformed back into the β-space (uniformity of density is preserved under linear transformations). From a sufficiently large sample of such variations of the power spectrum, we further select those functions for which successive peaks and dips are separated by at least 50 multipole moments l so that high-frequency wiggles are not counted as peaks or dips. Based on cosmological considerations, we also restricted the search to functions with 3 peaks (WMAP 1-, 3-, and 5-year data) or 4 peaks (7-year data). Peak and dip locations and heights were recorded for each peak and dip over this set of functions. This results into an empirical scatterplot that is indicative of the joint distribution of location and height for each peak or dip, under the assumption of uniform surface density on the confidence set $\mathcal{D}_{N,\alpha}$ [39].

Fig. 5 shows the results of probing the 95% confidence sets for uncertainties on peaks and dips, as outlined above (5000 acceptable function variations for each data realization). The box around a peak or a dip represents the largest horizontal and vertical variations in the scatter. In accordance with the confidence interval defined in Eq. 15, these form the 95% confidence intervals on the location and height of a peak/dip. Table II lists these confidence intervals together with 95% confidence intervals on peak height ratios.

As is well-known, the first peak was very clearly resolved in the 1-year data itself. Our results are manifestly consistent with this observation in the sense that its box is does not overlap with any other box. Further, our results clearly indicate that the second peak is resolved cleanly only in the 5-year data, whereas the third and fourth peaks are not resolved completely even in the 7-year data.

D. Uncertainties on the acoustic scale (l_A) and peak shift (ϕ_m) parameters

Consider the following relationship [31, 34] between the location l_m of the mth peak, the acoustic scale l_A, and the shift parameter ϕ_m:

$$l_m = l_A(m - \phi_m). \tag{19}$$

Substituting the end-points of the 95% confidence interval for the mth peak location, this relationship results into a hyperbolic band of allowed values in the $l_A - \phi_m$ plane. Such bands, derived from 95% confidence intervals on the first three peaks (Table II), are shown in Fig. 6. Additional information from other sources is required to constrain these bands to physically meaningful regions in the $l_A - \phi_m$ plane. For example, if we assume $l_A = 300$ [35] then, based on the 7-year data, the 95% confidence intervals for ϕ_m will be $\phi_1 : (0.1600, 0.3767), \phi_2 : (0.0367, 0.3600), \phi_3 : (-0.2167, 0.7300)$. Conversely, additional constraints on ϕ_m could be used to generate a confidence interval on l_A. From a model-independent point of view, we note that the (l_A, ϕ_m) bands for different peaks m appear to overlap around $\phi_m \approx 0$ and $200 \lesssim l_A \lesssim 400$. We interprete this as a nonparametric revelation of the nearly harmonic structure of peaks in the CMB power spectrum.

E. The low-l up-turn from a nonparametric viewpoint

Another interesting feature in Fig. 5 is the tiny but clearly observable scatter for the very first dip at the low-l end. This scatter corresponds to extreme power spectrum variations that reside on the surface of the 95% confidence set and have an up-turn at low l values. In the ΛCDM cosmology, such up-turn at the low-l end is primarily the result of the integrated Sachs-Wolfe (ISW) effect, and is seen in all parametric ΛCDM fits in Fig. 4. It would therefore be interesting to see what could be said about the low-l up-turn (and therby about the ISW effect) based on the nonparametric confidence set.

Notice that our nonparametric fits, which are at the centre of their respective confidence sets, do not show a low-l up-turn. However, the 7-year parametric ΛCDM fit, e.g., does show a clear up-turn at the low-l end. This parametric fit is at a distance corresponding to confidence level of about 10% ($\approx 0.12\sigma$; see Table I) from our 7-year nonparametric full-freedom monotone fit. This means that the confidence set for the 7-year nonparametric fit contains spectra with a low-l up-turn at most as far away as the 7-year parametric fit. We therefore conclude conservatively that the low-l up-turn as a feature of the CMB angular power spectrum cannot be ruled out at any confidence level in excess of about 10% [40].

IV. CONCLUSION

In this paper, we have presented a comparative nonparametric analysis of the WMAP 1-, 3-, 5-, and 7-year data releases for the CMB angular power spectrum, using a nonparametric function estimation methodology [1, 2]. In the context of this methodology, we have also presented our own numerical formulation for minimization of the inverse-noise-weighted risk function subject to monotonicity constraints, and a prescription for obtaining monotone nonparametric fits that are closer to cosmological expectations on smoothness. For all data realizations, we have presented results pertaining to the following questions: (a) how well is the angular power spectrum determined by the data alone, (b) how well is the ΛCDM model supported by a model-independent, nonparametric, data-driven analysis, and (c) what are the realistic uncertainties on peak/dip locations and heights.

The motivation for the analysis presented here was to explore what could be inferred about the CMB angular power spectrum in a model-independent, data-driven manner. On the other hand, the basic physics of the CMB is quite well established. It would therefore be useful to connect a nonparametric/model-independent analysis such as ours with the known physics of the CMB angular power spectrum. This is reserved for the future.

To conclude, we have demonstrated in this paper the threefold utility of the nonparametric methodology used here for cosmological function estimation problems: as a method with sound formal guarantees, as a sanity-enforcing mechanism on parametric model-based analyses, and as a method that allows interesting inferential questions to be addressed and answered in a data-driven manner.

Acknowledgments

MA is deeply indebted to Christopher R. Genovese and Larry Wasserman for many enlightening discussions covering all of statistics. Our `R` codes for computing the nonparametric fit are based on original codes by Christopher R. Genovese. TS would like to acknowledge support from the DST Swarnajayanti Fellowship.

[1] C. R. Genovese, C. J. Miller, R. C. Nichol, M. Arjunwadkar, and L. Wasserman, Statist. Sci. **19**, 308 (2004).

[2] B. Bryan, J. Schneider, C. J. Miller, R. C. Nichol, C. R. Genovese, and L. Wasserman, Astrophys. J. **665**, 25 (2007).
[3] G. Hinshaw, D. N. Spergel, L. Verde, R. S. Hill, S. S. Meyer, C. Barnes, C. L. Bennett, M. Halpern, N. Jarosik, A. Kogut, et al., Astrophys. J. Suppl. Ser. **148**, 135 (2003).
[4] G. Hinshaw, M. R. Nolta, C. L. Bennett, R. Bean, O. Doré, M. R. Greason, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, et al., Astrophys. J. Suppl. Ser. **170**, 288 (2007).
[5] M. R. Nolta, J. Dunkley, R. S. Hill, G. Hinshaw, E. Komatsu, D. Larson, L. Page, D. N. Spergel, C. L. Bennett, B. Gold, et al., Astrophys. J. Suppl. Ser. **180**, 296 (2009).
[6] D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, et al., Astrophys. J. Suppl. Ser. **192**, 16 (2011).
[7] C. L. Bennett, M. Bay, M. Halpern, G. Hinshaw, C. Jackson, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, et al., Astrophys. J. **583**, 1 (2003).
[8] J. A. Tauber, N. Mandolesi, J.-L. Puget, T. Banos, M. Bersanelli, F. R. Bouchet, R. C. Butler, J. Charra, G. Crone, J. Dodsworth, et al., Astron. Astrophys. **520**, A1 (2010).
[9] M. Tegmark, Phys. Rev. D **55**, 5895 (1997).
[10] G. Hinshaw, C. Barnes, C. L. Bennett, M. R. Greason, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, et al., Astrophys. J. Suppl. Ser. **148**, 63 (2003).
[11] G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, N. Jarosik, et al., Astrophys. J. Suppl. Ser. **180**, 225 (2009).
[12] N. Jarosik, C. Barnes, M. R. Greason, R. S. Hill, M. R. Nolta, N. Odegard, J. L. Weiland, R. Bean, C. L. Bennett, O. Doré, et al., Astrophys. J. Suppl. Ser. **170**, 263 (2007).
[13] N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, et al., Astrophys. J. Suppl. Ser. **192**, 14 (2011).
[14] R. Beran, J. Amer. Statist. Assoc. **95**, 155 (2000).
[15] R. Beran, in *Asymptotics in statistics and probability: papers in honor of George Gregory Roussas*, edited by M. L. Puri (VSP International Science Publishers, 2000), pp. 1–16.
[16] R. Beran and L. Dümbgen, Ann. Statist. **26**, 1826 (1998).
[17] R. Beran, Ann. Inst. Statist. Math. **48**, 1 (1996).
[18] L. Wasserman, *All of Nonparametric Statistics* (Springer-Verlag New York, 2006).
[19] P. I. C. A. The Pittsburgh Institute for Computational Astrostatistics, in *Statistical Challenges in Astronomy*, edited by E. D. Feigelson and G. J. Babu (Springer-Verlag New York, 2003), pp. 221–241, URL `http://arxiv.org/abs/astro-ph/0112050`.
[20] C. J. Miller, R. C. Nichol, C. R. Genovese, and L. Wasserman, Astrophys. J. **565**, L67 (2002).
[21] L. Verde, H. V. Peiris, D. N. Spergel, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, et al., Astrophys. J. Suppl. Ser. **148**, 195 (2003).
[22] T. Robertson, F. T. Wright, and R. L. Dykstra, *Order Restricted Statistical Inference* (John Wiley and Sons, 1988).
[23] B. N. Pshenichny and Y. M. Danilin, *Numerical Methods in Extremal Problems* (Mir Publishers, Moscow, 1978).
[24] M. J. D. Powell, in *Mathematical Programming Essays in Honor of George B. Dantzig Part II*, edited by R. W. Cottle, L. C. W. Dixon, B. Korte, T. L. Magnanti, M. J. Todd, E. L. Allgower, R. Bartels, V. Chvatal, J. E. Dennis, B. C. Eaves, et al. (Springer, Berlin, 1985), vol. 25 of *Mathematical Programming Studies*, pp. 46–61.
[25] D. Goldfarb and A. Idnani, Mathematical Programming **27**, 1 (1983).
[26] K. Schittkowski, Tech. Rep., University of Bayreuth, Department of Computer Science (2007), URL `http://www.klaus-schittkowski.de/`.
[27] R. J. Vanderbei, Optim. Methods Softw. **11**, 451 (1999).
[28] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2010), ISBN 3-900051-07-0, URL `http://www.R-project.org/`.
[29] J. R. Primack, J. Holtzman, A. Klypin, and D. O. Caldwell, Phys. Rev. Lett. **74**, 2160 (1995).
[30] J. R. Primack and M. A. K. Gross, in *Current Aspects of Neutrino Physics*, edited by D. O. Caldwell (Springer, Berlin, 2001), Physics and Astronomy Online Library, pp. 287–308.
[31] M. Doran and M. Lilley, Mon. Not. Roy. Astron. Soc. **330**, 965 (2002).
[32] R. Durrera, B. Novosyadlyj, and S. Apunevych, Astrophys. J. **583**, 33 (2003).
[33] D. E. Knuth, *Seminumerical Algorithms*, vol. 2 of *The Art of Computer Programming* (Addison-Wesley, 1981), 2nd ed.
[34] W. Hu, M. Fukugita, M. Zaldarriaga, and M. Tegmark, Astrophys. J. **549**, 669 (2001).
[35] L. Page, M. R. Nolta, C. Barnes, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, et al., Astrophys. J. Suppl. Ser. **148**, 233 (2003).
[36] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538**, 473 (2000), astro-ph/9911177.
[37] For a linear smoother, EDoF of a fit is formally defined as $\mathrm{tr}(H)$ [18], where H is the *hat matrix* that connects the fitted values $\widehat{Y}$ to the data Y, as $\widehat{Y} = HY$. For the present nonparametric regression method, $H = UDU^T$, where U is the orthonormal basis matrix (Eq. 5), and $D \equiv \mathrm{diag}(\lambda_0, \ldots, \lambda_{N-1})$. This implies $\mathrm{tr}(H) = \mathrm{tr}(D) = \sum_{i=0}^{N-1} \lambda_i$.
[38] In practice, the risk estimator (Eq. 9) and/or $\widehat{\tau}^2$ (Eq. 13) may turn out to be negative for particular data/covariance matrix realizations. In such cases, the squared confidence radius (Eq. 12) may be negative (or may not be 0 for $\alpha = 0$). For minimization purposes, the risk estimator (Eq. 9) is adequate and appropriate [16]. For confidence radius purposes, we suggest the following modifications to avoid the negativity problem:

$$\widehat{R}_+ = Z^T \bar{D} W \bar{D} Z + \max\left\{0, \mathrm{tr}(DWDB) - \mathrm{tr}(\bar{D} W \bar{D} B)\right\}$$

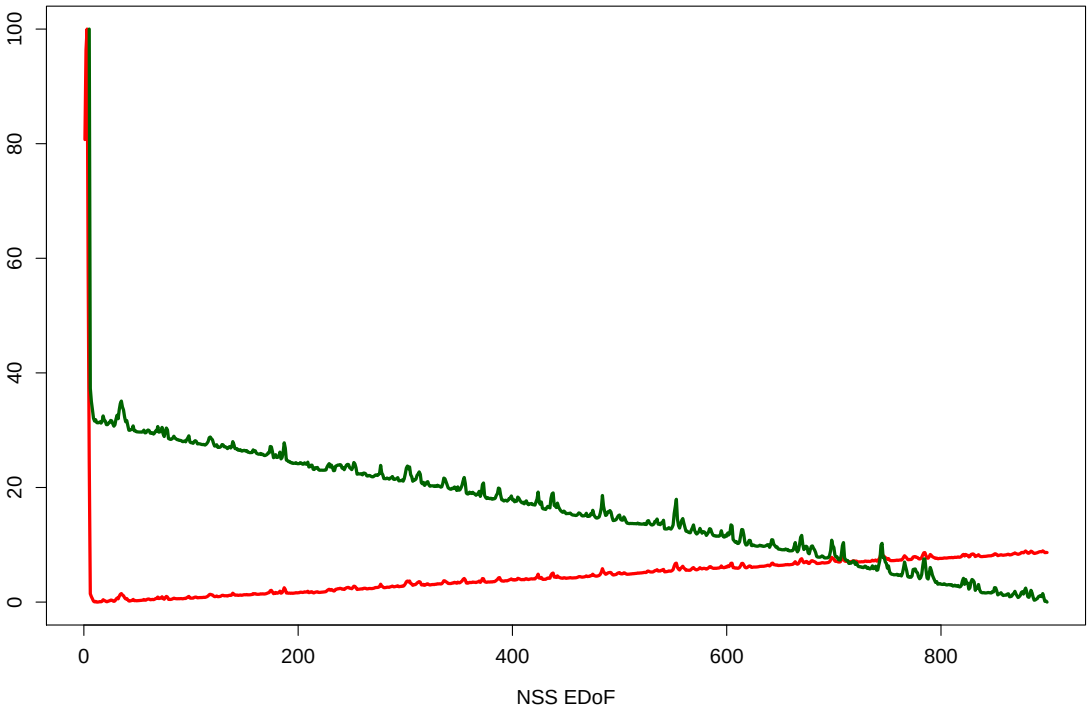

FIG. 1: Nonparametric risk (red curve) and −log(likelihood) (green curve) as functions of the EDoF for NSS fits (WMAP 1-year data), illustrating the contrasting behavior of the two quantities: Optimal smoothing occurs at EDoF = 12 where the nonparametric risk attains its minimum over the NSS set of fits. Likelihood function, on the other hand, keeps on improving with the EDoF indefinitely. Likelihood values are computed using the WMAP 1-year likelihood code [21]. The two quantities are rescaled to the common range $[0, 100]$ for display purposes.

$$\widehat{\tau}_+^2/N = 2\mathrm{tr}(ABAB) + \max\left\{0, Z^T QZ - \mathrm{tr}(QB)\right\}$$

$$r_{\alpha+}^2 = \max\left\{0, \frac{\widehat{\tau}_+ z_\alpha}{\sqrt{N}} + \widehat{R}_+\right\}.$$

At worst, this adjustment will make the confidence radius bigger, resulting into, e.g., wider confidence intervals, but more conservative inferences. Similar modifications have been suggested in [1, 14, 16].

[39] Note that the confidence set construct and the formal guarantees related to confidence intervals (Eq. 14) do not necessarily imply a uniform density over the confidence set. Uniform sampling was used as a convenient computational means of probing the confidence set surface in an unbiased fashion.

[40] Actually, there are indications that such up-turned variations of the spectrum may be much closer to the centre of the confidence set for the 7-year full-freedom monotone fit; this needs further investigation.

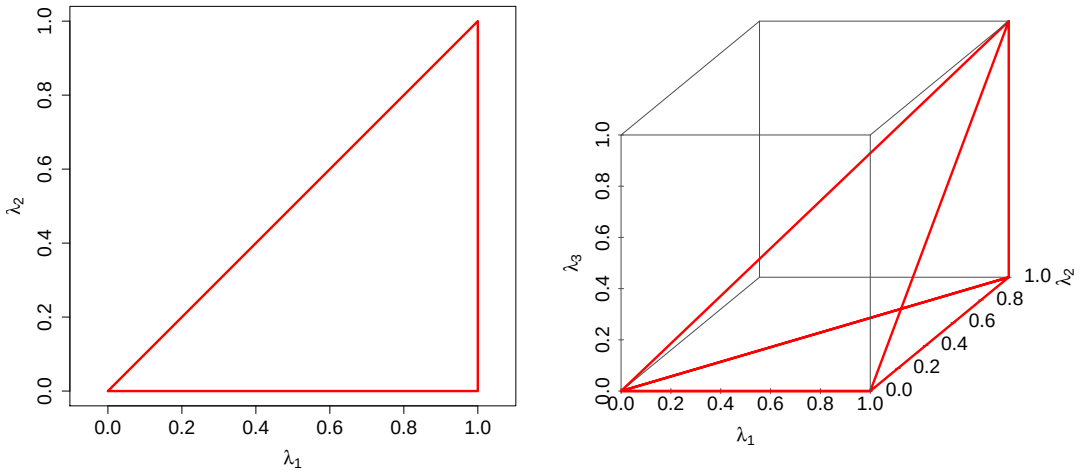

FIG. 2: Trianguloid-shaped admissible regions, marked by red lines, for the monotonicity constraint $1 \geq \lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{N-1} \geq 0$ for $N = 2$ and 3. The $(N+1)$ vertices of the trianguloid correspond to the $(N+1)$ NSS fits, with the origin corresponding to the zero function, and the vertex $(1, 1, \ldots, 1)$ corresponding to the function that exactly interpolates through the data. Surfaces with constant value p of EDoF are hyperplanes of the form $\sum_{i=0}^{N-1} \lambda_i = p$.

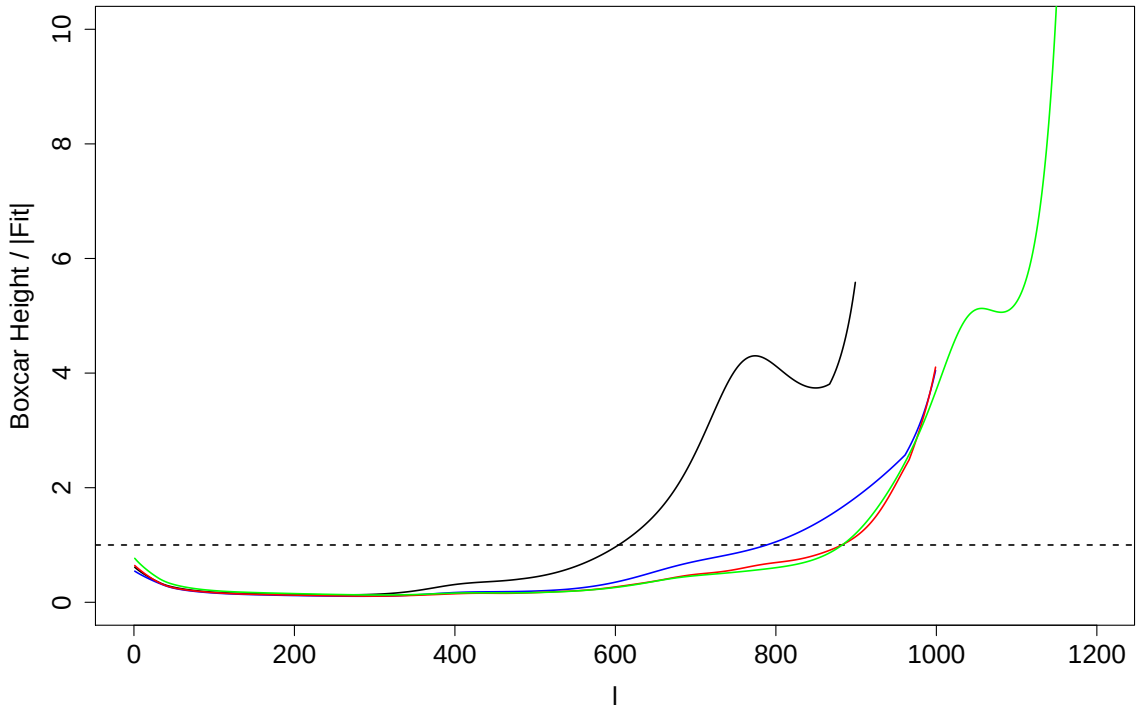

FIG. 3: The boxcar probe [1] for WMAP 1- (black), 3- (blue), 5- (red), 7-year (green) data. This is the maximum height of a flat localized additive perturbation to the fit so that the resulting spectrum remains in the 95% (2σ) confidence set for the fit, divided by the absolute value of the fit. This quantity is an approximate measure of how well the angular power spectrum is determined by the data: Values $\ll 1$ indicate that the fit is tightly determined by the data, whereas values > 1 indicate that the data contain little or no information about the height of the angular power spectrum for that l. Boxcar width was taken to be 60.

Data	ΛCDM	HΛCDM
1-year	0.1540	0.1493
$r_\alpha = 0.3679$	16.70%	15.87%
3-year	0.1462	0.2057
$r_\alpha = 0.3653$	14.52%	29.03%
5-year	0.1419	0.3552
$r_\alpha = 0.3563$	19.66%	94.82%
7-year	0.1238	0.3550
$r_\alpha = 0.3551$	9.08%	94.98%

TABLE I: Distances of model-based power spectra (ΛCDM and HΛCDM models; see Fig. 4 caption for details) from nonparametric full-freedom monotone fits. r_α is the confidence radius at $\alpha = 0.05$ (i.e., 95% confidence level $\equiv 2\sigma$). Percentages reported are the confidence levels corresponding to these distances; these can be interpreted as asymptotic probabilities with which the corresponding parametric fit is ruled out as a candidate for the true but unknown spectrum. Notice the dramatic progression of (a) how this HΛCDM model is pushed to the boundary of the confidence set, and (b) how the ΛCDM model gets closer to the nonparametric fit as the data become precise.

Data	Peak Location	Peak Height	Dip Location	Dip Height	Peak Location Ratio	Peak Height Ratio
1-year	$l_1: (186, 252)$	$h_1: (4968, 6133)$	$l_{1+\frac{1}{2}}: (378, 507)$	$h_{1+\frac{1}{2}}: (1152, 2029)$		
	$l_2: (440, 680)$	$h_2: (1879, 3933)$	$l_{2+\frac{1}{2}}: (494, 772)$	$h_{2+\frac{1}{2}}: (-2929, 2868)$	$l_2/l_1: (1.953, 3.217)$	$h_2/h_1: (0.329, 0.724)$
	$l_3: (559, 897)$	$h_3: (41, 10733)$	$l_{3+\frac{1}{2}}: (639, 900)$	$h_{3+\frac{1}{2}}: (-11376, 9962)$	$l_3/l_1: (2.495, 4.385)$	$h_3/h_1: (0.0075, 1.906)$
3-year	$l_1: (184, 254)$	$h_1: (5130, 6199)$	$l_{1+\frac{1}{2}}: (384, 456)$	$h_{1+\frac{1}{2}}: (1402, 1917)$		
	$l_2: (479, 606)$	$h_2: (2172, 2997)$	$l_{2+\frac{1}{2}}: (593, 856)$	$h_{2+\frac{1}{2}}: (80, 2145)$	$l_2/l_1: (2.028, 3.000)$	$h_2/h_1: (0.380, 0.555)$
	$l_3: (646, 977)$	$h_3: (1506, 6045)$	$l_{3+\frac{1}{2}}: (729, 1000)$	$h_{3+\frac{1}{2}}: (-4501, 5113)$	$l_3/l_1: (2.761, 4.801)$	$h_3/h_1: (0.264, 1.082)$
5-year	$l_1: (187, 251)$	$h_1: (5249, 6301)$	$l_{1+\frac{1}{2}}: (387, 445)$	$h_{1+\frac{1}{2}}: (1489, 1934)$		
	$l_2: (485, 596)$	$h_2: (2306, 2955)$	$l_{2+\frac{1}{2}}: (608, 793)$	$h_{2+\frac{1}{2}}: (971, 2095)$	$l_2/l_1: (2.040, 2.963)$	$h_2/h_1: (0.381, 0.525)$
	$l_3: (663, 982)$	$h_3: (1863, 4635)$	$l_{3+\frac{1}{2}}: (731, 1000)$	$h_{3+\frac{1}{2}}: (-2978, 3157)$	$l_3/l_1: (2.883, 4.672)$	$h_3/h_1: (0.321, 0.833)$
7-year	$l_1: (187, 252)$	$h_1: (5177, 6377)$	$l_{1+\frac{1}{2}}: (390, 442)$	$h_{1+\frac{1}{2}}: (1512, 1931)$		
	$l_2: (492, 589)$	$h_2: (2328, 3015)$	$l_{2+\frac{1}{2}}: (623, 803)$	$h_{2+\frac{1}{2}}: (1074, 2063)$	$l_2/l_1: (2.060, 2.887)$	$h_2/h_1: (0.386, 0.534)$
	$l_3: (681, 965)$	$h_3: (1871, 4119)$	$l_{3+\frac{1}{2}}: (733, 1104)$	$h_{3+\frac{1}{2}}: (-3111, 2709)$	$l_3/l_1: (3.000, 4.553)$	$h_3/h_1: (0.323, 0.732)$
	$l_4: (815, 1193)$	$h_4: (-1391, 7726)$	$l_{4+\frac{1}{2}}: (951, 1200)$	$h_{4+\frac{1}{2}}: (-11102, 6364)$	$l_4/l_1: (3.606, 6.047)$	$h_4/h_1: (-0.244, 1.340)$

TABLE II: 95% confidence intervals on several quantities directly related to features of the angular power spectrum. Negative values for heights and height ratios is a reflection of the high noise at the high-l end. Here, l_m (h_m) stands for the location (height) of the mth peak, and $l_{m+\frac{1}{2}}$ ($h_{m+\frac{1}{2}}$) denotes the location (depth) of the mth dip.

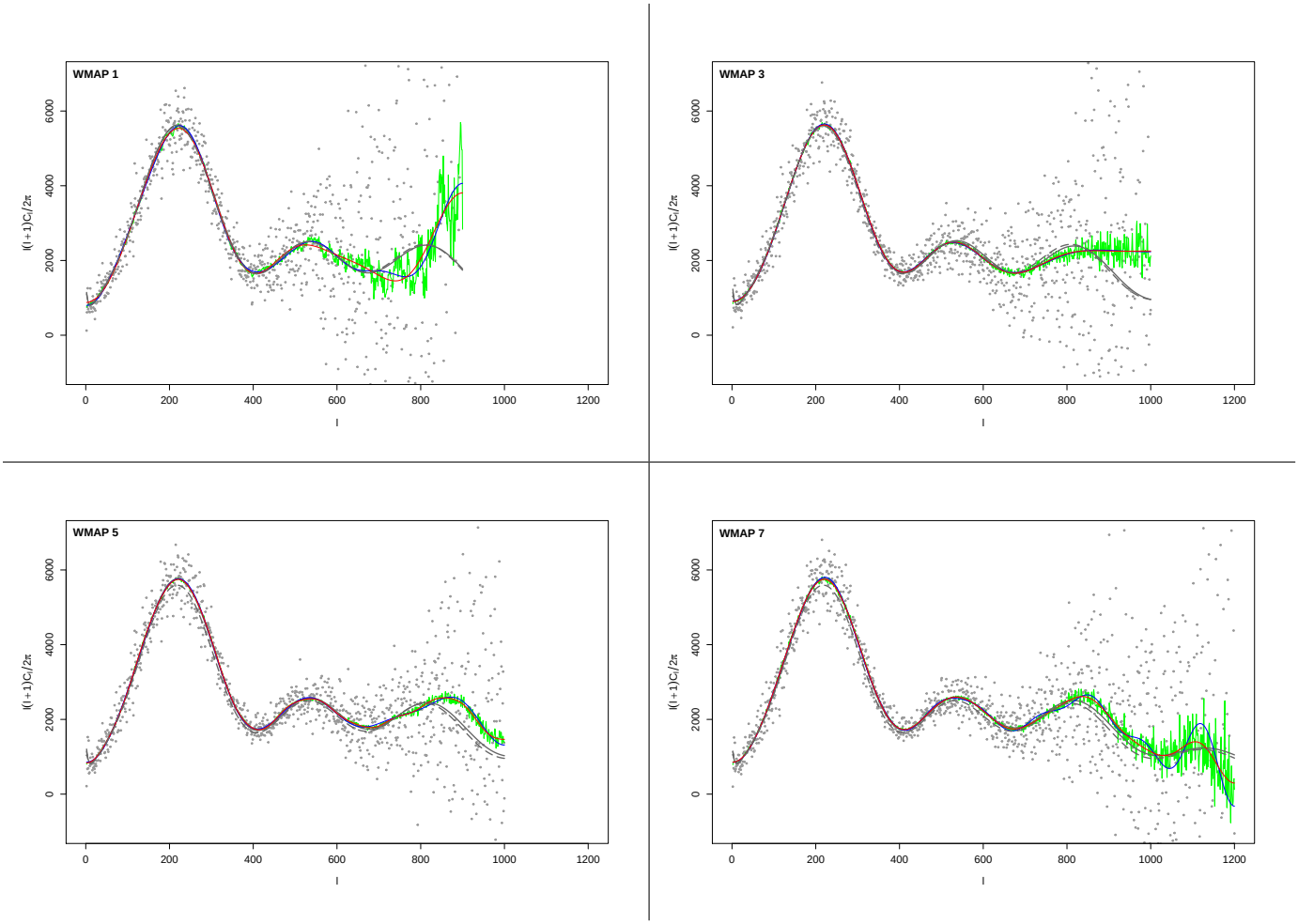

FIG. 4: Nonparametric fits for the four WMAP data sets. Notice that x and y ranges are identical across plots. Green: full-freedom monotone fit (EDoF $\approx 80.2, 76.5, 60.4, 102.9$ respectively); blue: NSS fit (EDoF $= 12, 10, 13, 20$ respectively); red: restricted-freedom monotone fit (EDoF $\approx 9.4, 9.5, 14.4, 14.1$ respectively); solid gray: best ΛCDM-based parametric fits [3–6] for the corresponding data realization; dashed gray: angular power spectrum for an HΛCDM model [29, 30]. The HΛCDM model considered here for illustrative purposes is defined by a small neutrino fraction ($\Omega_\nu h^2 = 0.00275$) with corresponding adjustment to the dark energy content ($\Omega_\Lambda = 0.729756$), and the rest of the parameters (including zero curvature) being identical to that of the best ΛCDM model [6] for the 7-year data. The HΛCDM angular power spectrum displayed was generated using the CAMB software [36].

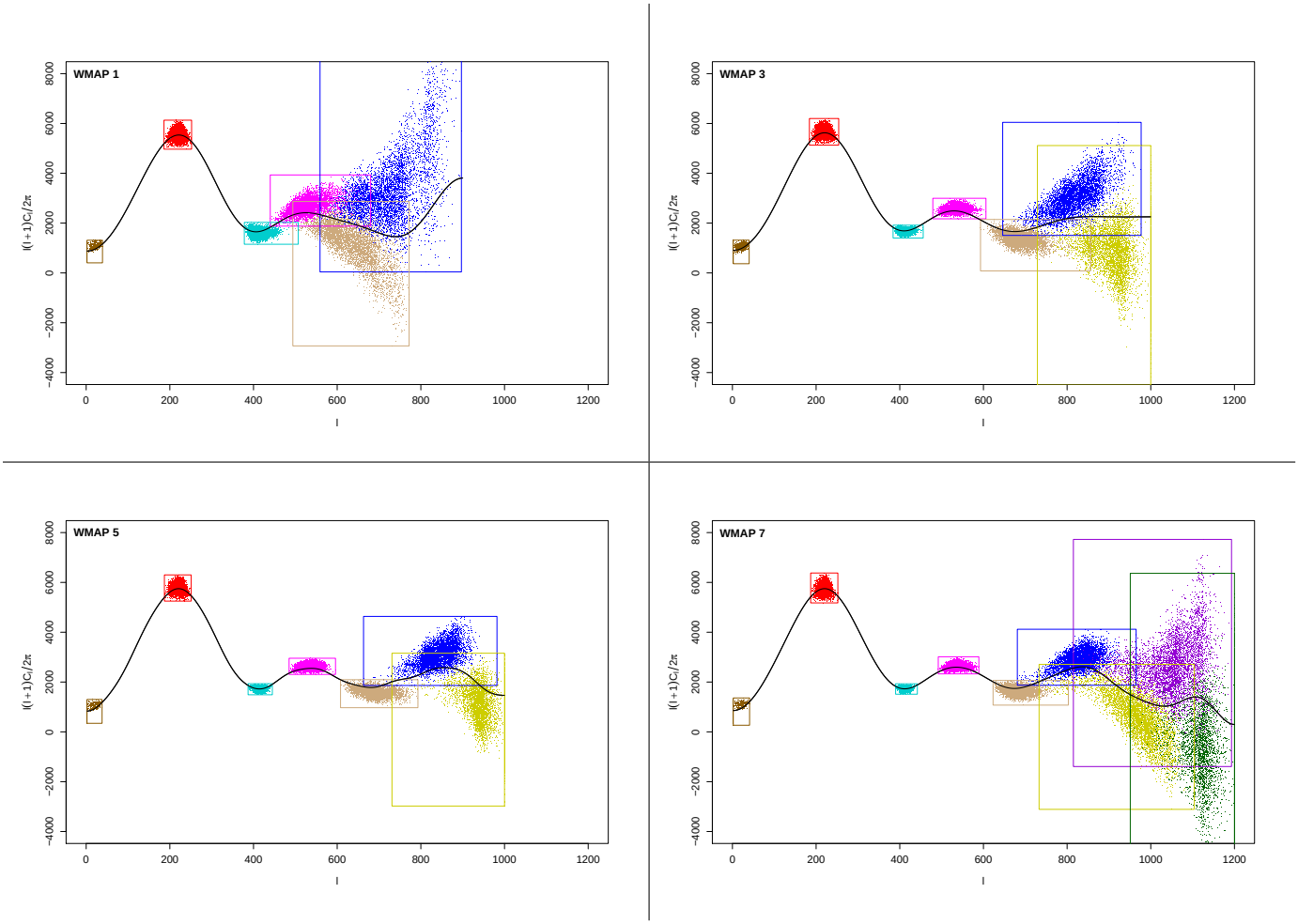

FIG. 5: Uncertainties on peak and dip locations and heights for the four WMAP data sets. Nonparametric fit displayed for reference is the restricted-freedom monotone fit in Fig. 4. The number of acceptable function variations sampled from the confidence set for each data realization is 5000.

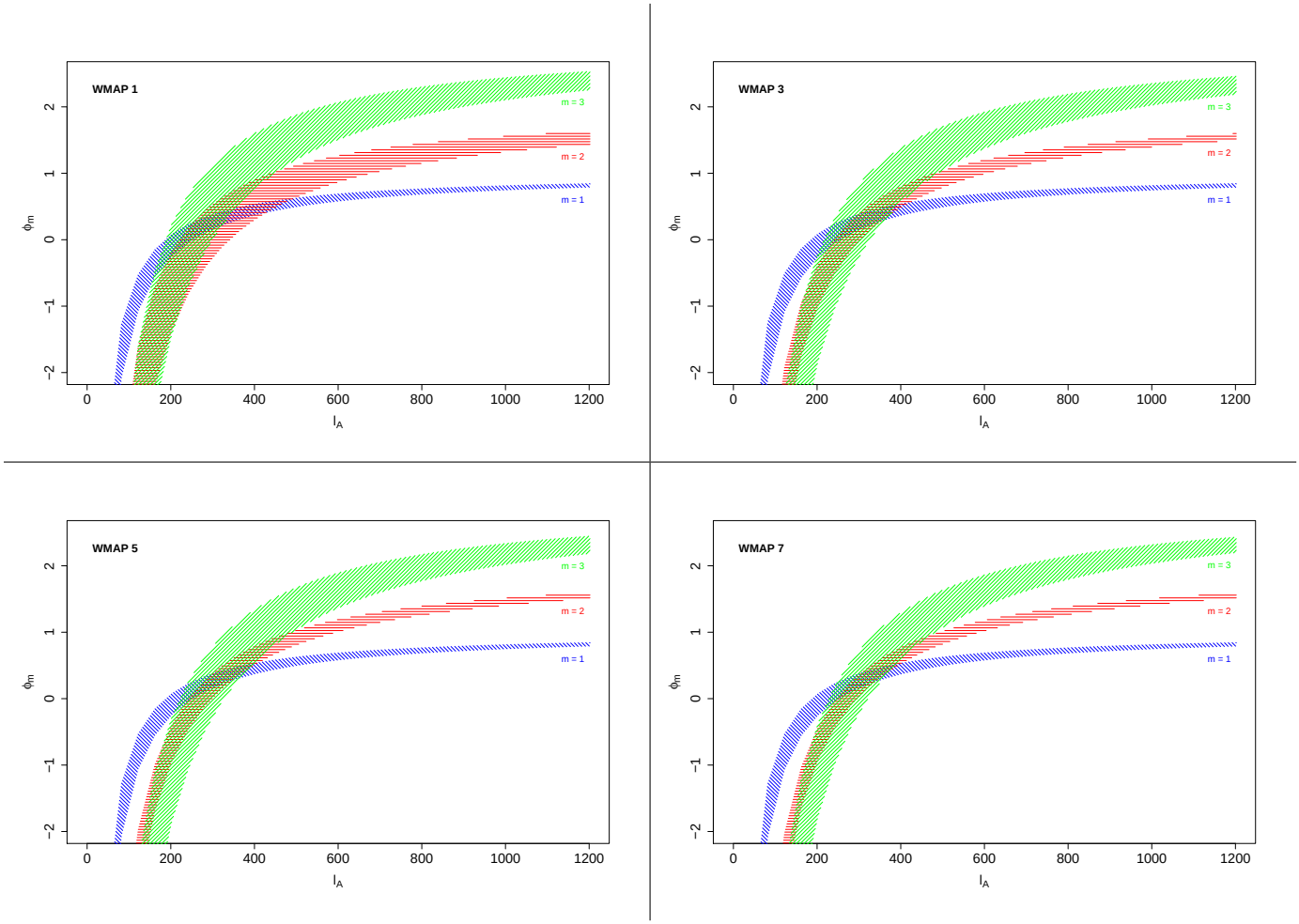

FIG. 6: Confidence "bands" for the acoustic scale l_A and the shift ϕ_m for the mth peak, as derived from the 95% confidence intervals on the first three peak locations (Table II) and Eq. 19. Blue: ϕ_1, red: ϕ_2, green: ϕ_3. Note that these (l_A, ϕ_m) bands for different peaks m appear to overlap around $\phi_m \approx 0$ and $200 \lesssim l_A \lesssim 400$: We interprete this as a nonparametric revelation of the nearly harmonic structure of peaks in the CMB power spectrum.