THE ASTROPHYSICAL JOURNAL, 439:L29–L32, 1995 February 1

GAUSSIAN STATISTICS OF THE COSMIC MICROWAVE BACKGROUND: CORRELATION OF TEMPERATURE EXTREMA IN THE *COBE*[1] DMR TWO-YEAR SKY MAPS

A. KOGUT,[2,3] A. J. BANDAY,[4] C. L. BENNETT,[5] G. HINSHAW,[2] P. M. LUBIN,[6] AND G. F. SMOOT[7]

Received 1994 August 22; accepted 1994 November 14

ABSTRACT

We use the two-point correlation function of the extrema points (peaks and valleys) in the *COBE* Differential Microwave Radiometers (DMR) 2 year sky maps as a test for non-Gaussian temperature distribution in the cosmic microwave background anisotropy. A maximum-likelihood analysis compares the DMR data to $n = 1$ toy models whose random-phase spherical harmonic components a_{lm} are drawn from either Gaussian, χ^2, or log-normal parent populations. The likelihood of the 53 GHz (A+B)/2 data is greatest for the exact Gaussian model. There is less than 10% chance that the non-Gaussian models tested describe the DMR data, limited primarily by type II errors in the statistical inference. The extrema correlation function is a stronger test for this class of non-Gaussian models than topological statistics such as the genus.

Subject headings: cosmic microwave background — methods: statistical

1. INTRODUCTION

The angular distribution of the cosmic microwave background (CMB) probes the distribution of mass and energy in the early universe and provides a means to test competing models of structure formation. One such test is whether or not the distribution of CMB anisotropies follows Gaussian statistics. In most inflationary models, the large-scale CMB anisotropy results from quantum fluctuations and follows their Gaussian statistics. Competing models (topological defects, axions, late phase transitions) generally involve higher order correlations and produce non-Gaussian distributions. Attempts to differentiate Gaussian from non-Gaussian distributions on large angular scales are complicated by the tendency of any distribution to approach Gaussian when averaged over a sufficiently large area (the central limit theorem) and by our inability to measure more than one sample (our observable universe) of the theoretical parent distribution ("cosmic variance").

Several authors (Hinshaw et al. 1994; Smoot et al. 1994; Luo 1994) have tested the first-year anisotropy maps from the *COBE* DMR experiment and find excellent statistical agreement with the hypothesis that the observed temperature fluctuations reflect random-phase Gaussian initial perturbations. However, since no competing models are examined, the compatibility with *non*-Gaussian models is not tested. In this *Letter*, we employ the two-point correlation function of extrema points to compare the two-year DMR maps to a set of broadly applicable toy models employing both Gaussian and non-Gaussian statistics. Simulations employing inputs with known distributions indicate that this statistic can successfully distinguish Gaussian from non-Gaussian toy models with $\sim 10\%$ error rate even at 10° angular resolution, and provide impetus for more computer-intensive studies of specific non-Gaussian cosmological models.

2. ANALYSIS

We test for Gaussian statistics using the set of extrema points in the temperature field $T(\theta, \phi)$, defined as those points for which $\nabla T = 0$. For a pixelized map, this reduces to the collection of pixels hotter or colder than all of their nearest neighbors. Specifying pixels hotter than their neighbors produces a set of "hot spots" or "peaks," while specifying colder pixels produces "cold spots" or "valleys." An additional data selection may be performed, requiring $|T|$ to be greater than some threshold ν, usually expressed in terms of the standard deviation σ of the temperature field.

The 2 point correlation function of the extrema pixels provides a compact description of the data,

$$C_{\rm ext}(\theta) = \frac{\sum_{i,j} w_i w_j T_i T_j}{\sum_{i,j} w_i w_j},$$

where w is some weighting factor and the sum runs over all pixel pairs $\{i, j\}$ separated by angle θ. We consider three applications of the extrema correlation function: peak-peak (autocorrelation of just the peaks or just the valleys), peak-valley (cross-correlation of the peak pixels with the valley pixels), and combined extrema (autocorrelation of all extrema points without regard for their second derivative). Bond & Efstathiou (1987) provide analytic approximations for these functions for random Gaussian fields but do not explicitly include the effects of instrument noise superposed on the CMB. Since the correlation properties of the nonuniform noise in the DMR maps are different from the underlying CMB temperature field, we use Monte Carlo techniques instead to derive the mean extrema correlation function and covariance as a function of the threshold ν.

We analyze the extrema correlation functions of the 2 year *COBE* DMR maps (Bennett et al. 1994) and compare the sensitive 53 GHz (A+B)/2 sum maps and (A−B)/2 difference maps

[1] The National Aeronautics and Space Administration/Goddard Space Flight Center (NASA/GSFC) is responsible for the design, development, and operation of the *Cosmic Background Explorer* (*COBE*). Scientific guidance is provided by the *COBE* Science Working Group. GSFC is also responsible for the development of analysis software and for the production of the mission data sets.

[2] Hughes STX Corporation, Laboratory for Astronomy and Solar Physics, Code 685, NASA/GSFC, Greenbelt, MD 20771.

[3] kogut@stars.gsfc.nasa.gov.

[4] Universities Space Research Association, Laboratory for Astronomy and Solar Physics, Code 685, NASA/GSFC, Greenbelt, MD 20771.

[5] Laboratory for Astronomy and Solar Physics, Code 685, NASA Goddard Space Flight Center, Greenbelt, MD 20771.

[6] UCSB Physics Department, Santa Barbara, CA 93106.

[7] LBL, SSL, & CfPA, Bldg 50-351, University of California, Berkeley, CA 94720.

to Monte Carlo simulations of scale-invariant ($n = 1$) CMB anisotropy superposed with instrument noise. We generate each CMB realization using a spherical harmonic decomposition $T(\theta, \phi) = \sum_{lm} a_{lm} Y_{lm}(\theta, \phi)$ in which the harmonic coefficients a_{lm} are random variables with zero mean and l-dependent variance,

$$\langle a_{lm}^2 \rangle = (Q_{\rm rms-PS})^2 \frac{4\pi}{5} \frac{\Gamma[l + (n-1)/2]\Gamma[(9-n)/2]}{\Gamma[l + (5-n)/2]\Gamma[(3+n)/2]}$$

(Bond & Efstathiou 1987), where $Q_{\rm rms-PS}$ is the power-spectrum amplitude normalization expressed at the quadrupole (Smoot et al. 1992). The coefficients a_{lm} are drawn from parent populations with either Gaussian, log-normal, or χ_N^2 ($N = 1$, 5, or 15 degrees of freedom) distributions, normalized to the mean and variances above. A non-Gaussian amplitude distribution for the a_{lm} while retaining random phases provides a simple modification to the standard Gaussian model of CMB anisotropy. The log-normal distribution is the most strongly non-Gaussian, while the χ_N^2 models provide a smooth transition from strongly non-Gaussian ($N = 1$) to nearly Gaussian ($N = 15$). The models tested are not an exhaustive set of non-Gaussian models but are a computationally simple test of the power of various statistics on large angular scales. Although the non-Gaussian *amplitude* distributions tested here are skew-positive, the resulting sky maps are the convolution of the a_{lm} with the spherical harmonics Y_{lm} and are thus characterized by a positive kurtosis in the distribution of temperatures T (e.g., higher "wings" than a Gaussian distribution). We test the sensitivity of our results to the transformation $a_{lm} \rightarrow -a_{lm}$ and find no difference using either definition.

The coefficients a_{lm} define toy models to which specific models of structure formation may be compared (e.g., Weinberg & Cole 1992). On large angular scales, non-Gaussian models such as cosmic strings or textures approximate the scale-invariant power spectrum of the standard Gaussian inflationary models. On smaller angular scales strongly non-Gaussian features appear, either as phase correlations (line defects in string models) or as an overabundance of high-amplitude hot and cold spots (texture models). Our toy models do not introduce phase correlations, but the increased probability to produce high-amplitude fluctuations in the positive-kurtosis χ^2 or log-normal models qualitatively reproduces this feature of textures or other models with rare high-amplitude peaks. The toy models cannot be compared in detail to specific physically motivated non-Gaussian models, but share enough of their characteristics to motivate further investigation of these CPU-intensive models if the toy models demonstrate reasonable discrimination versus the standard Gaussian model.

We generate 1000 $n = 1$ full-sky realizations for each CMB model. To each CMB realization we add a realization of instrument noise defined by the level and pattern of noise in the DMR 2 year 53A and 53B channels (Bennett et al. 1994), then combine the channels to form (A+B)/2 sum maps and (A−B)/2 difference maps. We do not include Galactic emission or systematic uncertainties since these are small compared to the noise (Bennett et al. 1992, 1994; Kogut et al. 1992). We smooth the maps with a 7° Gaussian full width at half-maximum as a compromise between suppressing noise and removing power at small scales, resulting in an effective smoothing on the sky of 10°. We reject pixels with Galactic latitude $|b| < 20°$, remove fitted monopole and dipole temperatures from the surviving pixels, and determine the standard deviation σ. A nearest-neighbor algorithm then forms the collection of extrema pixels at thresholds $\nu = [0, 1, 2]\sigma$; these pixels at each threshold are then used to generate the peak-peak, peak-valley, and combined extrema correlation functions using unit weighting and 2°.6 bins in the separation angle θ. Since, by definition, two peaks cannot be adjacent, we ignore the bin at zero separation and the first nonzero bin in all subsequent analysis. Analysis shows that the results are dominated by the first few remaining bins; consequently, we speed processing by truncating the correlation function at separation $\theta = 60°$ for a total of 22 angular bins.

The correlation function $C_{\rm ext}|_\nu$ evaluated at thresholds $\nu = [0, 1, 2]\sigma$ defines a vector $S = [C_{\rm ext}|_{\nu=0}, C_{\rm ext}|_{\nu=1}, C_{\rm ext}|_{\nu=2}]$, consisting of $k = 66$ components (three sets of 22 angular bins). The properties of each toy model may be specified by the mean value of this vector in each bin i, $\langle S_i \rangle = (1/N) \sum S_i$, and the covariance matrix $M_{ij} = (1/N) \sum (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle)$ between bins i and j derived from the N realizations of the vector S. We compare the DMR data to the simulations via a Gaussian approximation to the likelihood

$$\mathcal{L}(Q_{\rm rms-PS}, \Upsilon) = (2\pi)^{-k/2} \frac{\exp[-(1/2)\chi^2]}{\sqrt{\det(M)}},$$

where $\chi^2 = \sum_{ij} (D_i - \langle S_i \rangle)(M^{-1})_{ij}(D_j - \langle S_j \rangle)$, and D is a similar vector of correlation functions for the DMR maps. The likelihood $\mathcal{L}$ is a function of two parameters: the continuous variable $Q_{\rm rms-PS}$ representing the normalization and the discrete variable Υ representing the five a_{lm} toy models. Since the covariance matrix M depends strongly on these parameters, a simple χ^2 minimization approach fails, leading us to use the likelihood instead as a statistical tool in the Monte Carlo simulations described below. We evaluate the likelihood $\mathcal{L}$ in the 2-dimensional parameter space for values of $Q_{\rm rms-PS}$ spanning the range [0, 30] μK and search for the maximum in the resulting distribution.

3. RESULTS AND DISCUSSIONS

Figure 1 shows the likelihood of the 2 year 53 GHz (A+B)/2 maps for both the Gaussian and the non-Gaussian models derived from the autocorrelation function of all extrema pixels, normalized to unity for the most likely model. The likelihood function is greatest for the exact Gaussian model, with relative likelihoods for the non-Gaussian models ~ 0.05. Restricting the analysis to the $n = 1$ Gaussian model yields a maximum likelihood normalization for the 53 GHz (A+B)/2 maps of $Q_{\rm rms-PS} = 18.1 \pm 1.9$ μK, in agreement with other estimates of $Q_{\rm rms-PS}$ using the 2 year *COBE* data (Wright et al. 1994; Górski et al. 1994; Bennett et al. 1994; Banday et al. 1994). The width of the distribution in $Q_{\rm rms-PS}$ is similar for all models. The likelihood function of the (A−B)/2 difference maps peaks at $Q_{\rm rms-PS} = 0$ with no significant preference between models. Similar results occur for the peak-peak and peak-valley correlation functions.

The small likelihoods in Figure 1 for the non-Gaussian toy models given the DMR data would seem to rule out these models at high statistical confidence. However, formal identification of confidence intervals relies heavily on assumptions of the statistical distributions in the analysis (e.g., that the residuals $D - \langle S \rangle$ are multivariate normal) which are not always realized in practice. Furthermore, since the parameter Υ represents a collection of discrete models instead of a con-

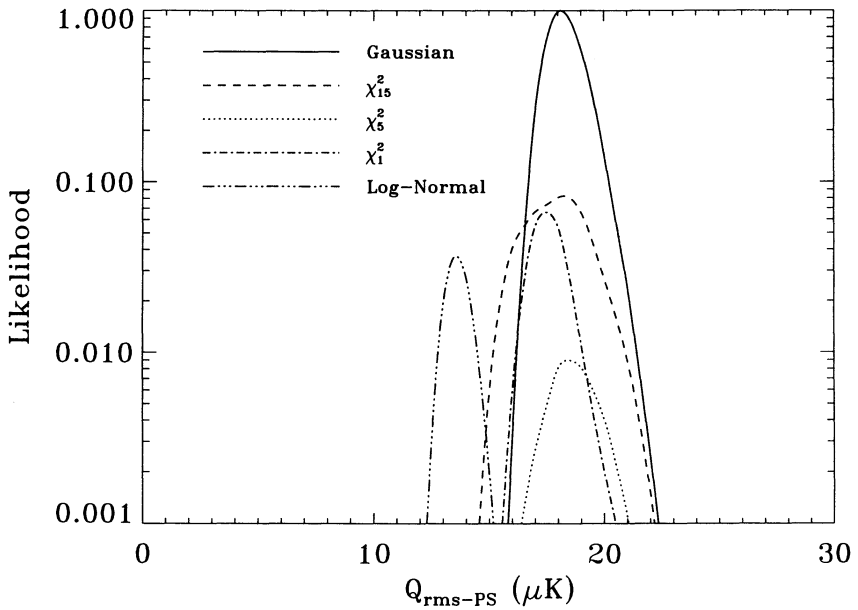

FIG. 1.—Likelihood function of the 2 year DMR 53 GHz (A + B)/2 extrema correlation function for Gaussian and non-Gaussian toy models

tinuous variable, we cannot integrate over Υ to derive confidence intervals in the usual way. Although the degree of freedom N in the χ^2_N distribution could in principle serve as an integration variable, we prefer a more general approach which will remain viable for discrete physically motivated models such as topological defects for which no such variable exists. We resolve these problems, assign relative probabilities, and test for statistical bias using a Monte Carlo approach.

We generate 1000 realizations of one model (e.g., Gaussian) with $Q_{rms-PS} = 18$ μK and use the same machinery (likelihood analysis of the extrema correlation function) to generate the likelihood $\mathcal{L}(Q_{rms-PS}, \Upsilon)$ of each realization against all five toy models, replacing the DMR correlation vector D with the vector S from each realization in turn. We select the likelihood maximum in Q_{rms-PS} for each model Υ (effectively ignoring any information on Q_{rms-PS}) and study the resulting set of likelihood maxima to see how often the largest likelihood occurs at each model Υ. This is equivalent to generating Figure 1 for each realization, evaluating only the relative heights of each likelihood curve, and counting how often each model is selected as "most likely." We then generate another 1000 realizations with a different input model and repeat for all five toy models.

Table 1 shows the percentage of simulations for which each model was selected as "most likely" as a function of the model used to generate the realizations. When the input is known to be Gaussian (col. [2]), 61% of the simulations correctly identify the exact Gaussian as the "best" model, with the remainder incorrectly allocated among the non-Gaussian models (a type I error). When the input is instead one of the non-Gaussian toy models (cols. [3]–[6]), that model is correctly identified in a similar fraction of the realizations (note that the χ^2_1 and log-normal distributions are nearly degenerate). There is no evidence for any statistical bias favoring one particular model: the fact that the DMR data "prefer" the Gaussian model is not an artifact of the method.

Given that the DMR likelihood is greatest for the exact Gaussian model, how confident are we that the CMB is *not* in fact a realization of one of the non-Gaussian toy models (a type II error)? From the first row of Table 1 (Gaussian model received highest likelihood) we see that the probability of obtaining this result is 3 times larger for a Gaussian CMB than for any of the non-Gaussian toy models. There is a 64% relative probability that the CMB follows Gaussian statistics and only 36% probability that it is better described by one of the non-Gaussian toy models.

A more powerful test uses additional information from the likelihood distribution. We have examined the subset of simulations for which the best-fitted model was not, in fact, the correct input, and found that the likelihoods in these cases did not strongly select against the rejected models. The DMR likelihood does not show this pattern: the second-best likelihood (for the χ^2_{15} model) is only 0.08. Table 2 shows the percentage of simulations for which each model was selected as "most likely" while the next-best likelihood was smaller than 0.08. We recover the same overall pattern as Table 1: the most probable outcome is to recover the input model correctly, but the fraction of both type I errors (columns) and type II errors (rows) is reduced. The relative probability to obtain the DMR result is 5–10 times greater for the Gaussian CMB model than the non-Gaussian toy models: there is less than 10%–20% chance that any of these models describes the DMR data.

TABLE 1

PERCENTAGE OF SIMULATIONS FROM SINGLE TEST

FITTED MODEL	INPUT MODEL[a]				
	GAUSSIAN	χ^2_{15}	χ^2_5	χ^2_1	LOG-NORMAL
Gaussian.....	61	21	20	23	24
χ^2_{15}.....	16	56	21	21	23
χ^2_5.....	11	11	44	15	18
χ^2_1.....	5	6	7	28	15
log-normal.....	7	6	8	13	20

[a] Percentage of 1000 simulations for which each model was selected as "most likely" as a function of the model used to generate the realizations.

TABLE 2

PERCENTAGE OF SIMULATIONS FROM DOUBLE TEST

FITTED MODEL	INPUT MODEL[a]				
	GAUSSIAN	χ^2_{15}	χ^2_5	χ^2_1	LOG-NORMAL
Gaussian.....	20	2	2	3	4
χ^2_{15}.....	1	16	1	3	3
χ^2_5.....	1	1	10	1	2
χ^2_1.....	2	2	3	15	8
log-normal.....	4	4	5	7	13

[a] Percentage of 1000 simulations for which each model was selected as "most likely" with next-best likelihood less than 0.08.

The topological quantity known as the genus has also been proposed as a test for non-Gaussian statistics in the CMB (Gott et al. 1990). Smoot et al. (1994) show that the genus of the first-year DMR maps is consistent with random-phase Gaussian models. We have performed a likelihood analysis of the genus of the 2 year DMR maps compared to the same set of non-Gaussian toy models used for the extrema correlation analysis above. Although the genus likelihood is also greatest for the exact Gaussian model, the ability to reject the non-Gaussian models is weaker, with maximum-likelihood Max ($\mathscr{L}$) ≈ 0.3 for the log-normal and χ^2_1 models using the genus compared to Max ($\mathscr{L}$) ≈ 0.05 using the extrema correlation function. Coles & Barrow (1987) discuss the genus of somewhat different χ^2 models and reach a similar conclusion that the genus does not strongly differentiate between random-phase Gaussian and non-Gaussian models. The genus of the 2 year DMR maps will be discussed in greater detail in a future paper.

Both the genus and the extrema correlation function show the 2 year DMR data to be consistent with the hypothesis of random-phase Gaussian statistics and inconsistent with random-phase toy models with non-Gaussian distributions of the spherical harmonic coefficients a_{lm}. There is less than 10% probability that the non-Gaussian models tested describe the large angular scale anisotropy of the CMB. Although the statistical power of these tests is not overwhelming, they do demonstrate that large-beam experiments can probe the statistical distribution of CMB anisotropy. Physically motivated non-Gaussian models (e.g., topological defects) have strong phase correlations as well, which would be expected to increase the statistical power of these tests. There is thus an incentive to pursue further tests of specific models using the *COBE* DMR data.

We gratefully acknowledge the dedicated efforts of those responsible for the *COBE* DMR data. C. Lawrence, C. Lineweaver, and L. Tenorio provided helpful discussion of statistical techniques. *COBE* is supported by the Office of Space Sciences of NASA Headquarters.

REFERENCES

Banday, A. J., et al. 1994, ApJ, 436, 99
Bennett, C. L., et al. 1992, ApJ, 396, L7
———. 1994, ApJ, 436, 423
Bond, J. R., & Efstathiou, G. 1987, MNRAS, 226, 655
Coles, P., & Barrow, J. D. 1987, MNRAS, 228, 407
Górski, K. M., Hinshaw, G., Banday, A. J., Bennett, C. L., Wright, E. L., Kogut, A., Smoot, G. F., & Lubin, P. 1994, 430, L89
Gott, J. R., Park, C., Juszkiewicz, R., Bies, W. E., Bennett, D. P., Bouchet, F. R., & Stebbins, A. 1990, ApJ, 352, 1
Hinshaw, G., et al. 1994, ApJ, 431, 1
Kogut, A., et al. 1992, ApJ, 401, 1
Luo, X. 1994, Phys. Rev. D, 49, 3810
Smoot, G. F., Tenorio, L., Banday, A. J., Kogut, A., Wright, E. L., Hinshaw, G., & Bennett, C. L. 1994, ApJ, 437, 283
Smoot, G. F., et al. 1992, ApJ, 396, L1
Weinberg, D. H., & Cole, S. 1992, MNRAS, 259, 652
Wright, E. L., Smoot, G. F., Bennett, C. L., & Lubin, P. M. 1994, ApJ, 436, 443