

TESTS FOR NON-GAUSSIAN STATISTICS IN THE DMR FOUR-YEAR SKY MAPS

A. KOGUT,^{1,2} A. J. BANDAY,^{1,3} C. L. BENNETT,⁴ K. M. GÓRSKI,^{1,5} G. HINSHAW,¹ G. F. SMOOT,⁶ AND E. L. WRIGHT⁷

Received 1996 January 9; accepted 1996 March 22

ABSTRACT

We search the high-latitude portion of the *COBE* Differential Microwave Radiometer (DMR) 4 yr sky maps for evidence of a non-Gaussian temperature distribution in the cosmic microwave background. The genus, three-point correlation function, and two-point correlation function of temperature maxima and minima are all in excellent agreement with the hypothesis that the CMB anisotropy on angular scales of 7° or larger represents a random-phase Gaussian field. A likelihood comparison of the DMR sky maps to a set of random-phase non-Gaussian toy models selects the exact Gaussian model as most likely. Monte Carlo simulations show that the two-point correlation of the peaks and valleys in the maps provides the greatest discrimination among the class of models tested.

Subject headings: cosmic microwave background — cosmology: observations — methods: statistical

1. INTRODUCTION

The angular distribution of the cosmic microwave background (CMB) reflects the density distribution at the epoch of last scattering. The statistical distribution of CMB temperature fluctuations is thus an important probe of the initial conditions for structure formation.

Causally connected regions at the surface of last scattering, as viewed from the present epoch, subtend an angle $\sim 2^\circ$. Most of the current models for structure formation predict the CMB temperature fluctuations to be nearly Gaussian on larger angular scales. In inflationary cosmologies, this Gaussian distribution reflects the quantum origin of the density perturbations, while in topological defect models, it results from the superposition of a large number of non-Gaussian perturbations at smaller angular scales (the central limit theorem).

The CMB temperature field may be conveniently represented in terms of spherical harmonics,

$$T(\theta, \phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta, \phi) \quad (1)$$

and is said to be Gaussian if the distribution each coefficient $a_{\ell m}$ follows a Gaussian probability distribution with phases uniformly distributed in the range $[0, 2\pi]$. Gaussian fields have the desirable property that they are completely specified by the power spectrum or its Fourier transform, the two-point correlation function. Thus, if the CMB is demonstrated to be a Gaussian field, not only have current models passed an important consistency test, but we are also spared the task of tabulating the full set of higher order correlation functions. Deviations from Gaussian statistics would necessitate a dramatic reappraisal of current models.

The *COBE*⁸ Differential Microwave Radiometer (DMR) experiment mapped the CMB at 7° angular resolution. A number of authors have tested the DMR 1 and 2 yr sky maps and found good agreement with a Gaussian distribution (Smoot et al. 1994; Luo 1994; Hinshaw et al. 1994, 1995; Torres et al. 1995). Kogut et al. (1995) compared the 2 yr DMR maps to non-Gaussian toy models and found that the exact Gaussian model was the most likely descriptor of the data. In this Letter we compare the final 4 yr DMR maps (Bennett et al. 1996; Kogut et al. 1996) to both Gaussian and non-Gaussian toy models and discuss the relative ability of various tests to discriminate between the Gaussian and non-Gaussian models tested.

2. COMPARISON WITH GAUSSIAN MODELS

Cosmic variance, sample variance, and instrument noise can combine to create non-Gaussian realizations from initially Gaussian parent populations, while the central limit theorem pushes the CMB distribution back to Gaussian on large angular scales (Scherrer & Schaefer 1995). We infer the statistical properties of the CMB parent population using Monte Carlo techniques, in which a certain statistic (three-point correlation function, extrema correlation function, or genus) is calculated for many simulations of noisy CMB skies. The simulations include all relevant aspects of the CMB model, instrument noise, and data handling, including removal of the Galactic plane and fitted dipole/quadrupole subtraction. The same statistic is then calculated for the DMR data and compared to the range observed in the simulations. If the DMR data fall comfortably within the range of equivalent Monte Carlo realizations, we accept the model as in agreement with the data; otherwise, we reject the model at some specified confidence level.

Unless otherwise noted, we model the CMB as a scale-invariant Gaussian field [$P(k) \propto Q_{\text{rms-PS}}^2 k^n$ with $n = 1$] with quadrupole power-spectrum normalization $Q_{\text{rms-PS}} = 18 \mu\text{K}$ (Górski et al. 1996), evaluated at multipoles $2 \leq \ell \leq 40$ filtered through the DMR window function (Wright et al.

¹ Hughes STX Corporation, Laboratory for Astronomy and Solar Physics, Code 685, NASA/GSFC, Greenbelt, MD 20771.

² kogut@stars.gsfc.nasa.gov.

³ Current address: Max-Planck-Institut für Astrophysik, Karl-Swarzschild-Strasse 1, 85740 Garching bei München, Germany.

⁴ Laboratory for Astronomy and Solar Physics, Code 685, NASA/GSFC, Greenbelt, MD 20771.

⁵ On leave from Warsaw University Observatory, Aleje Ujazdowskie 4, 00-478 Warsaw, Poland.

⁶ Department of Physics, Lawrence Berkeley Laboratory, Space Sciences Laboratory, and Center for Particle Astrophysics, Building 50-351, University of California, Berkeley, CA 94720.

⁷ UCLA Astronomy, P.O. Box 951562, Los Angeles, CA 90095-1562.

⁸ 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 analysis software and for the production of the mission data sets.

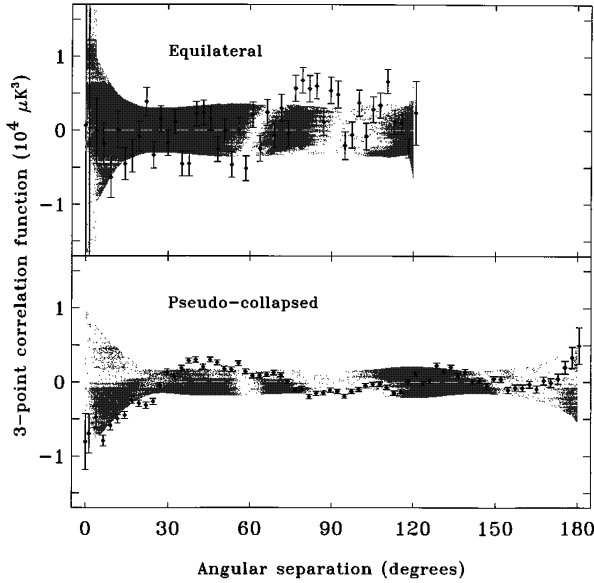


FIG. 1.—Three-point correlation function of the 4 yr 53 GHz $(A + B)/2$ sum map, compared to simulated maps with $Q_{\text{rms-PS}} = 18 \mu\text{K}$, $n = 1$, and 53 GHz noise properties (thermodynamic temperature units). The error bars represent the uncertainty due to instrument noise alone, while the gray bands represent the 68% confidence region from the combined noise and Gaussian sky signal. *Top*: Equilateral triangle configuration. *Bottom*: Pseudocollapsed configuration. The fluctuations about zero correlation are too large to be ascribed to instrument noise alone but are consistent with the range of fluctuations expected from a Gaussian CMB.

1994). We compare these simulations to the DMR 4 yr 53 GHz $(A + B)/2$ sum map. All statistics are evaluated only for the high-latitude sky (3881 pixels at $|b| > 20^\circ$ with custom cutouts near Ophiuchus and Orion; Bennett et al. 1996) from which a fitted dipole and quadrupole have been removed.

2.1. Three-Point Correlation Function

A Gaussian field is fully characterized by its two-point correlation function. All odd moments are identically zero, while all higher even moments may be specified in terms of the two-point function. The three-point correlation function is a first test for non-Gaussian signatures in the DMR data. The general three-point correlation function is the average product of three temperatures with a fixed relative orientation on the sky. For computational reasons we evaluate only two special configurations: the equilateral configuration, in which the three legs form an equilateral triangle, and the more sensitive pseudocollapsed configuration, in which two of the legs are nearly parallel (for exact definitions, see Hinshaw et al. 1994). Figure 1 shows the equilateral and pseudocollapsed three-point correlation functions for the 4 yr DMR 53 GHz $(A + B)/2$ maps, evaluated with unit pixel weight in 2.6° angular bins. The gray band in each plot shows the 68% confidence region derived from 2000 random realizations of scale-invariant CMB anisotropy superposed with instrument noise and includes the effects of cosmic and sample variance. Deviations from zero for either configuration are well within the range expected for a single realization of a stochastic process. We quantify this using a χ^2 statistic,

$$\chi^2 = \sum_{\alpha\beta} (D_\alpha - \langle S_\alpha \rangle) (\mathbf{M}^{-1})_{\alpha\beta} (D_\beta - \langle S_\beta \rangle), \quad (2)$$

where D_α is the DMR three-point correlation function, $\langle S_\alpha \rangle$ is the mean correlation function from the simulations (in this case identically zero), and

$$\mathbf{M}_{\alpha\beta} = \frac{1}{N} \sum_i (S_\alpha^{(i)} - \langle S_\alpha \rangle) (S_\beta^{(i)} - \langle S_\beta \rangle) \quad (3)$$

is the mean covariance matrix between the angular bins α and β computed from the $N = 2000$ correlation functions $S^{(i)}$ in the simulations. We calculate 2000 such χ^2 values by replacing the three-point correlation function from the DMR data, D_α , with the correlation function from the i th simulation, $S_\alpha^{(i)}$. The DMR maps fall near the median of the resulting distribution: 66% of the simulations had larger χ^2 than the DMR pseudocollapsed three-point correlation function, while 31% of the simulations had larger χ^2 than the DMR equilateral configuration. The three-point correlation function shows no evidence for a non-Gaussian CMB distribution.

2.2. Genus

Topology provides another test of the Gaussian hypothesis. The statistical properties of the CMB can be characterized by the excursion regions enclosed by isothermality contours. The genus is the total curvature of the contours at fixed temperature and may be loosely defined as the number of isolated high-temperature regions (hot spots) minus the number of isolated low-temperature regions (cold spots). For a random Gaussian field, the genus per unit area is

$$g(\nu) = \frac{1}{(2\pi)^{3/2} \theta_c^2} \nu e^{-\nu^2/2}$$

(Gott et al. 1990), where ν is the temperature contour threshold in units of the standard deviation σ of the field, and θ_c is the coherence angle, related to the two-point correlation function $C_2(\theta)$ by

$$\theta_c^{-2} = -\frac{1}{C_2(0)} \left. \frac{d^2 C_2}{d\theta^2} \right|_{\theta=0}.$$

The genus per unit area is a locally invariant quantity and is insensitive to incomplete sky coverage (e.g., removal of the Galactic plane).

The genus of a Gaussian field depends on two parameters: the coherence angle θ_c and the standard deviation σ . The DMR noise is nearly uncorrelated (Lineweaver et al. 1994); its coherence angle is determined by the pixelization. The coherence angle of a scale-free CMB depends only on the index n . For a noisy map, however, both the coherence angle and the standard deviation depend on the relative amplitude of the signal and noise terms and hence acquire a dependence on the CMB amplitude normalization $Q_{\text{rms-PS}}$. Since $\theta_{c,\text{noise}} \ll \theta_{c,\text{CMB}}$, the sensitivity of the genus to the CMB terms can be enhanced by smoothing the map, provided the smoothing dispersion remains less than $\theta_{c,\text{CMB}}$. In the unsmoothed DMR maps, the correlation angles are 1.4° and 9.3° for the instrument noise and CMB anisotropy, respectively. We evaluate the genus at three smoothing angles with full width at half-maximum 0° (no smoothing), 5° , and 10° . At each smoothing angle, a nearest-neighbor algorithm counts the number of hot and cold spots (contiguous pixels above or below the temperature threshold) at 11 thresholds ν from -2.5σ to $+2.5\sigma$ in steps of 0.5σ . We adopt the difference in spot number as an

estimate of the genus and use the Monte Carlo simulations to calibrate any difference induced by the Galactic cut between this definition and the total curvature. The two definitions are identical in the absence of a Galactic cut, and the nearest-neighbor algorithm allows significant computational savings.

As with the three-point correlation function, we compare the DMR 53 GHz map to Gaussian simulations using a χ^2 statistic (eqs. [2] and [3] with vectors D_α and S_α now consisting of the genus in $k = 33$ bins for the 11 thresholds and three smoothing angles evaluated simultaneously). The DMR genus is near the median of the Gaussian models: 51% of the simulations had larger χ^2 .

Since the instrument noise properties are known, we may interpret the genus in terms of the CMB model parameters. We evaluate the χ^2 of the DMR genus on a two-dimensional grid in $Q_{\text{rms-PS}}$ and n , use the value at each grid point to determine the likelihood

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

and search for the maximum in the resulting likelihood distribution. We also perform a probabilistic search by converting the χ^2 at each grid point to a probability $P(\chi^2)$, defined as the fraction of simulations with χ^2 larger than the DMR value at that grid point. We calibrate both methods and search for biases by generating 2000 simulations with fixed parameters $Q_{\text{rms-PS}}$ and n , fitting each realization, and examining the “cloud” of fitted values. The χ^2 probability test has a small bias of $0.8 \mu\text{K}$ toward smaller values of $Q_{\text{rms-PS}}$; the maximum likelihood test is unbiased.

Both methods yield consistent results on the DMR 53 GHz $(A + B)/2$ data. The likelihood shows a strong ridge along the line $Q_{\text{rms-PS}} = (17.3 \pm 1.8) - (8.8 \pm 0.2)(n - 1)$, μK with likelihood maximum at $Q_{\text{rms-PS}} = 24.5 \pm 5.0 \mu\text{K}$ and $n = 0.2 \pm 0.6$ (68% confidence interval). The value at the maximum has $P(\chi^2) = 77\%$, and is not statistically preferred over the more precise value derived from the power spectrum of the combined DMR maps (Górski et al. 1996).

2.3. Extrema Correlation Function

Another test of the statistical distribution is the angular correlation of the extrema (peaks and valleys) in the temperature field, 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 “peaks,” while specifying colder pixels produces “valleys.”

The number density of peaks or valleys is dominated by the noise properties of the map (Kogut et al. 1994). The clustering of the extrema, as measured by the two-point correlation function of the maxima and minima, provides additional information on the underlying CMB temperature field. We define the extrema correlation function as

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

where w_i are the pixel weights and the sum over pixel temperatures T_i is restricted to the set of maxima and minima pixels (Kogut et al. 1995). Since the correlation properties of the nonuniform noise in the DMR maps are different from the underlying CMB temperature field, we smooth the maps with

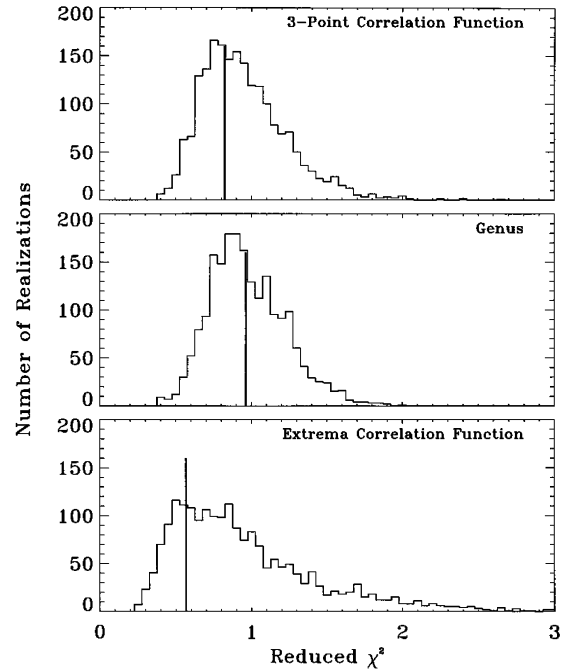


FIG. 2.— χ^2 of the DMR maps compared to the distribution derived from simulations with the Gaussian CMB model, divided by the number of bins used to evaluate the χ^2 . *Top*: Three-point correlation function. *Middle*: Genus. *Bottom*: Extrema correlation function. The χ^2 values from the DMR data (vertical bars) are near the median of the distribution expected from Gaussian CMB anisotropy.

a 7° FWHM Gaussian prior to collating the extrema as a compromise between suppressing noise and removing CMB power at small scales. We use Monte Carlo techniques as described above to calibrate the resulting distribution of simulated correlation functions.

The clustering of extrema depends on the threshold ν . We evaluate $C_{\text{ext}}(\theta)$ at thresholds $|\nu| = 0, 1$, and 2 . Since, by definition, two peaks cannot be adjacent, we ignore both the bin at zero angular separation and the first nonzero bin in all subsequent analysis. Simulations show that the results are dominated by the first few remaining bins; consequently, we speed processing by truncating the correlation function at separation $\theta = 60^\circ$ for a total of 22 angular bins at each of the three thresholds.

The angular distribution of peaks and valleys in the DMR maps is in excellent agreement with Gaussian statistics. We calculate a χ^2 value for the DMR extrema correlation function and compare it to 2000 simulations of CMB models with $Q_{\text{rms-PS}} = 18 \mu\text{K}$ and $n = 1$. Seventy percent of the simulations had larger χ^2 than the DMR maps.

3. COMPARISON WITH NON-GAUSSIAN MODELS

Figure 2 summarizes the comparison of the 53 GHz $(A + B)/2$ 4 yr sky maps with Gaussian CMB models. The χ^2 derived from the three-point correlation function, genus, and extrema correlation function of the DMR data fall near the median of the χ^2 distributions derived from the simulations: the DMR data are in excellent agreement with the hypothesis that the CMB is a realization of a parent population of Gaussian density perturbations.

Given the large DMR beam, the agreement of the DMR data with Gaussian models does not necessarily place tight

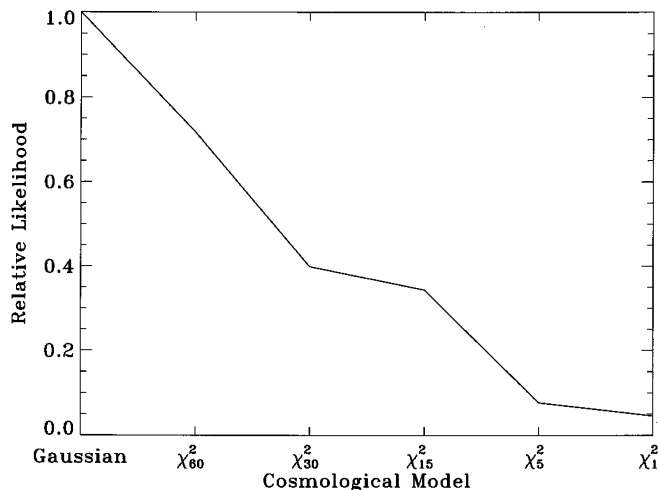


FIG. 3.—Likelihood function of the 4 yr DMR 53 GHz $(A + B)/2$ extrema correlation function (marginalized over the CMB normalization $Q_{\text{rms-PS}}$) for Gaussian and non-Gaussian models.

limits on *non*-Gaussian models. Ideally, we would use Monte Carlo realizations of specific cosmological models such as texture or global monopoles. Unfortunately, no analytic expressions exist for the temperature anisotropy produced by these models; generating a single realization requires a substantial computational investment to follow the causal field ordering in some large volume. Instead, we will modify the standard Gaussian model to test generic but interesting non-Gaussian amplitude distributions.

We use a likelihood analysis of the statistical tests described above to compare the DMR maps against a set of toy models for which the spherical harmonic amplitudes $a_{\ell m}$ (eq. [1]) are random variables with zero mean drawn from χ^2_N parent populations with $N = 1, 5, 15, 30$, or 60 degrees of freedom. We fix the variance σ_ℓ^2 of the parent population at the same value as the standard Gaussian model,

$$\sigma_\ell^2 = (Q_{\text{rms-PS}})^2 \frac{4\pi}{5} \frac{\Gamma[l + (n - 1)/2] \Gamma[9 - n]/2]}{\Gamma[l + (5 - n)/2] \Gamma[(3 + n)/2]}$$

(Bond & Efstathiou 1987), providing a simple, computationally efficient modification to the standard Gaussian model. The toy models provide a smooth transition from strongly non-Gaussian ($N = 1$) to nearly Gaussian ($N = 60$) and may be compared to specific models of structure formation (Weinberg & Cole 1992; Kogut et al. 1995). Cosmological models with rare high-amplitude peaks, typified by topological defect models such as texture or global monopoles can be compared to the χ^2 models with few degrees of freedom, which tend to produce such features.

Figure 3 shows the likelihood of the 4 yr DMR maps for the Gaussian and χ^2_N models, derived from the extrema correlation function. The likelihood is greatest for the exact Gaussian model and decreases as the toy models become increasingly non-Gaussian. The Gaussian model remains the most likely descriptor of the data when the genus replaces the extrema correlation in the likelihood analysis.

Given that the DMR data “prefer” the Gaussian model, how confident are we that the CMB is not well described by one of the χ^2 toy models? Since cosmic variance can produce a nearly Gaussian sky from a non-Gaussian parent population

and vice versa, marginalizing over model parameters can produce misleading confidence intervals (see the discussion of this point in Kogut et al. 1995). We quantify the ability of each statistical test to discriminate between Gaussian and non-Gaussian models using Monte Carlo simulations. We take 2000 simulations from each of the models in turn and for each simulation generate the likelihood function $\mathcal{L}(Q_{\text{rms-PS}}, N)$. We then count the number of times the Gaussian model is selected as “most likely.” We obtain the DMR result (Gaussian model chosen as most likely) 5 times more often when the input model is Gaussian than when the input is any of the non-Gaussian models. The method is unbiased: if we ask how often the χ^2_{15} model would be selected as most likely, we find the greatest probability if the input model is χ^2_{15} , and so forth. In this posterior sense, the CMB is 5 times more likely to be a realization of a Gaussian process than any of the χ^2 models tested. The extrema correlation function discriminates among the toy models twice as well as the genus.

4. CONCLUSIONS

We have tested the statistical distribution of temperature fluctuations in the DMR 4 yr 53 GHz $(A + B)/2$ summed sky map against both Gaussian and non-Gaussian models of CMB anisotropy using three statistical tests: the three-point correlation function, genus, and two-point correlation function of temperature maxima and minima. A goodness-of-fit test using the χ^2 of each function shows that the DMR data lies near the median of the distribution of χ^2 values evaluated for randomly generated Gaussian models with amplitude normalization $Q_{\text{rms-PS}} = 18 \mu\text{K}$ and power-law index $n = 1$. The DMR data are thus in excellent agreement with the hypothesis that the observed CMB anisotropy represents a realization of a Gaussian parent process. We test an alternate hypothesis, that the CMB represents a realization of a non-Gaussian process, by evaluating the likelihood of the DMR data against a grid of non-Gaussian toy models whose spherical harmonic coefficients $a_{\ell m}$ are drawn from χ^2_N distributions with $N = 1, 5, 15, 30$, or 60 degrees of freedom. The likelihood reaches a maximum for the exact Gaussian model. Cosmic variance produces significant cross talk among the class of models tested and precludes the simple identification of confidence intervals on N . Simulations in which a similar likelihood analysis is performed on random realizations of either Gaussian or non-Gaussian skies demonstrate that the method correctly identifies the input model with a $\sim 15\%$ error rate against any of the other models; the error rate is nearly independent of N . The DMR result (Gaussian model selected as most likely) occurs 5 times more often when the input to the simulations is, in fact, Gaussian, than when it is any of the non-Gaussian toy models. We thus conclude that not only do Gaussian power-law models provide adequate description of the CMB anisotropy on large angular scales, but that the non-Gaussian models tested are 5 times less likely to describe the true statistical distribution.

We gratefully acknowledge the dedicated efforts of the many people responsible for the *COBE* DMR data: the NASA office of Space Sciences, the *COBE* flight team, and all of those who helped process and analyze the data. This work was funded in part by NASA grant S-57778-F.

REFERENCES

- Bennett, C. L., et al. 1996, ApJ, 464, L1
Bond, J. R., & Efstathiou, G. 1987, MNRAS, 226, 655
Górski, K. M., Banday, A. J., Bennett, C. L., Hinshaw, G., Kogut, A., Smoot, G. F., & Wright, E. L. 1996, ApJ, 464, L11
Gott, J. R., Park, C., Juskiewicz, R., Biew, W. E., Bennett, D. P., Bochet, F. R., & Stebbins, A. 1990, ApJ, 352, 1
Hinshaw, G., et al. 1994, ApJ, 431, 1
Hinshaw, G., Banday, A. J., Bennett, C. L., Górski, K. M., & Kogut, A. 1995, ApJ, 446, L67
Kogut, A., Banday, A. J., Bennett, C. L., Górski, K. M., Hinshaw, G., Smoot, G. F., & Wright, E. L. 1996, 464, L5
Kogut, A., Banday, A. J., Bennett, C. L., Hinshaw, G., Loewenstein, K., Lubin, P., Smoot, G. F., & Wright, E. L. 1994, ApJ, 433, 435
Kogut, A., Banday, A. J., Bennett, C. L., Hinshaw, G., Lubin, P. M., & Smoot, G. F. 1995, ApJ, 439, L29
Lineweaver, C. H., et al. 1994, ApJ, 436, 452
Luo, X. 1994, Phys. Rev. D, 49, 3810
Scherrer, R. J., & Schaefer, R. K. 1995, ApJ, 446, 44
Smoot, G. F., Tenorio, L., Banday, A. J., Kogut, A., Wright, E. L., Hinshaw, G., & Bennett, C. L. 1994, ApJ, 437, 1
Torres, S., Cayón, L., Martínez-González, E., & Sanz, J. L. 1995, MNRAS, 274, 853
Weinberg, D. H., & Cole, S. 1992, MNRAS, 259, 652
Wright, E. L., et al. 1994, ApJ, 420, 1