# Statistical Significance of the Cobe Quadrupole Detection

Philip B. Stark

Technical Report No. 380
December 1992

Department of Statistics
University of California
Berkeley, California 94720

# STATISTICAL SIGNIFICANCE OF THE *COBE* QUADRUPOLE DETECTION

PHILIP B. STARK

Department of Statistics

University of California, Berkeley

December 18, 1992

## ABSTRACT

The uncertainties of estimates of cosmic microwave background (CMB) heterogeneity using Cosmic Background Explorer (*COBE*) data are extremely sensitive to assumptions about the CMB, especially where foreground sources dominate. At 4.5% statistical significance, the *COBE* least-squares estimate of the CMB quadrupole using data at galactic latitudes $|b| > 20°$ is consistent with a model in which the true quadrupole is zero, and multipoles of degrees 3-20 follow the Harrison-Zel'dovich ($n = 1$) spectrum normalized to the *COBE* $16\mu K$ quadrupole estimate derived from the power spectrum. Most of the uncertainty is due to possible CMB heterogeneity not modelled in the *COBE* data analysis. Physical constraints on the neglected high frequencies are needed to establish that the *COBE* estimate is significantly different from zero.

*Subject heading:* cosmic microwave background

## 1. INTRODUCTION

The Cosmic Background Explorer (*COBE*) is designed to measure long-wavelength cosmic microwave background (CMB) heterogeneity, which may discriminate among competing cosmological theories (see, *e.g.*, Wright et al. 1992). The *COBE* group found from their first year of data (3.4GB) that the CMB is heterogeneous, with rms quadrupole $Q_{rms} \approx 13 \pm 4\mu K$ based on least-squares fits, and $Q_{rms-PS} \approx 16 \pm 4\mu K$ using higher order moments of the power spectrum, assuming the spectrum of heterogeneity follows a power law (Smoot et al. 1992; Bennett et al. 1992 discuss galactic effects; Kogut et al. 1992 treat systematic errors). (Hogan (1992) and Wickramasinghe (1992) propose alternative physical explanations for the observed variations.) The purpose of this article is to show that the uncertainty neglected in approximating CMB structure by a simple parametric model is large enough to make the *COBE* quadrupole estimate statistically indistinguishable from zero, and that *a priori* constraints on the spatial spectrum of CMB fluctuations are needed to reduce the error bars on $Q_{rms}$ and individual multipole coefficients to interesting levels.

## 2. *A PRIORI* CONSTRAINTS ARE NECESSARY

Sampling of the CMB by *COBE* is uneven due to the orbit and spin of the instrument; in addition, data are censored when the instrument is pointed too close to the Earth or Moon. In the *COBE* data analysis it is necessary to omit data in a range of galactic latitudes $|b| < x$ to reduce contamination from foreground sources, so it is possible to construct a perturbation to the CMB that is undetectable in the censored data, but has a nonzero multipole coefficient. Adding a large multiple of that perturbation to any proposed CMB model yields a model that fits the data just as well, but has an arbitrarily different value of the multipole coefficient: multipole coefficients can not be determined by (even noise-free) data alone. Formally, the uncertainty in multipole coefficients is infinite in the absence of additional constraints. This observation is basic to inverse theory (Backus and Gilbert 1968).

If data from the entire sky were retained, the uncertainty in multipole coefficients would still be infinite without additional constraints, due to the fact that spherical harmonics are not linear combinations of the antenna beam patterns. There would still be perturbations with nonzero multipole coefficients, but orthogonal to all the measurement functionals.

For estimates of CMB multipoles to have finite uncertainty, we need constraints that rule out large, "invisible" perturbations. The Harrison-Zel'dovich model (see, *e.g.* Wright et al. 1992) in itself does not rule out large perturbations, since Gaussian processes are expected to have occasional large deviations. It might be that for an invisible perturbation to have a significant multipole coefficient it would have to be as large as the galactic signal. The next section shows that small perturbations can change the estimated quadrupole quite a bit.

## 3. SIMULATED UNCERTAINTY OF $Q_{rms}$ ESTIMATES

The principal procedure used to estimate the CMB quadrupole is to fit a small number of spherical harmonics to the data, either in original form or reduced to "sky maps" consisting of about 6000 pixels (Bennett et al. 1992; Kogut et al., 1992; Smoot et al. 1992). To show that physically reasonable perturbations can have substantial effects on least-squares $Q_{rms}$ estimates, quadrupoles were fit to data from 1000 pseudo-random CMB models using routines similar to the *COBE* routines. The models were generated according to the scale-invariant Harrison-Zel'dovich "$n = 1$ power law" (see, *e.g.*, Wright et al. 1992) in which multipole coefficients $a_{lm}$ are independent zero-mean Gaussian random variables, with variances $\sigma_l^2$ depending only on the degree $l$:

$$\sigma_l^2 = \frac{6Q^2_{rms-PS}}{5l(l+1)}. \tag{1}$$

$Q_{rms-PS}$ was taken to be the $16\mu K$ *COBE* estimate based on the power spectrum, which is likely to represent the power over this range of $l$ better than the least-squares $Q_{rms}$ estimates. The multipole coefficients were set to zero for $l < 3$ and $l > 20$, and no noise was added. The simulations were conditioned to have rms $\sqrt{2l+1}\sigma_l$ in each degree. (This is a way to account for "cosmic variance" due to the small fraction of the surface of last scattering in Earth's light cone.) As in the preferred *COBE* solution, data for galactic latitudes $|b| < 20°$ were omitted, and weights accounted for the uneven sampling elsewhere. The *COBE* team has also published solutions deleting data for $|b| < 10°$ and $|b| < 30°$ (Bennett *et al.*, 1992; Smoot *et al.*, 1992); omitting data for $|b| < 20°$ appears to be a reasonable compromise between the possibility of galactic contamination and throwing away most of the data.

Least-squares estimates of $Q_{rms}$ when the true value was zero ranged from $1.2\mu K$ to $15.6\mu K$, with 25th percentile, median, and 75th percentile equal to $4.8\mu K$, $6.5\mu K$ and $8.3\mu K$, respectively: in this model, "aliasing" of higher frequencies accounts for half the *COBE* $Q_{rms}$ estimate half of the time, even without noise, and even under the optimistic $n = 1$ power law. The true uncertainty is likely to be larger since the simulations neglect $l < 3$ and $l > 20$, and since the $n = 1$ power law might not hold. Assuming the $n = 1$ power law holds, one could find analytically the distribution of the contribution of neglected terms to the estimated quadrupole; the computation requires inverting the data mapping matrix.

## 4. STATISTICAL SIGNIFICANCE OF THE *COBE* QUADRUPOLE

The published uncertainty in the $|b| > 20°$ least-squares estimate of $Q_{rms}$ is $\pm 5\mu K$ at 68% confidence (Smoot et al. 1992). The figure $\pm 5\mu K$ includes stochastic and systematic contributions, but not the aliasing effect described above. Let $X$ denote the error giving rise to the $\pm 5\mu K$ uncertainty in Smoot et al. (1992). If we assume the probability distribution of $X$ is symmetric, the chance that $X \geq 5\mu K$ is $(1 - 0.68)/2 = 0.16$. Let $Y$ denote the aliasing error described above. Using the independence of $X$ and $Y$, we can lower-bound the probability that their sum exceeds the *COBE* least-squares estimate $Q_{rms} = 13\mu K$:

$$\begin{aligned} P(X+Y \geq 13) &\geq P(X \geq 5 \text{ and } Y \geq 8) & (2) \\ &= P(X \geq 5)P(Y \geq 8) \\ &\approx (0.16) * (0.28) = 0.045. & (3) \end{aligned}$$

That is, under this model, the $p$-value of the hypothesis that the true quadrupole is zero is at least 0.045: the *COBE* quadrupole estimate is at best barely statistically significant at level 0.05. The bound in inequality (2) is crude, so the true $p$-value is larger (the results are less significant). Furthermore, it is optimistic to assume *a priori* that $n = 1$. Thus while the *COBE* evidence for heterogeneity of some kind is strong, the uncertainty in the inferred $Q_{rms}$ is such that the least-squares $Q_{rms}$ estimate from data with $|b| > 20°$ is consistent with zero at about the 5% significance level. This does not mean that the *COBE* estimate is in error, just that there is a different hypothesis ($Q_{rms} = 0$) that is also consistent with the data within the uncertainties.

## 5. DISCUSSION

*Mitigating Effects.* The presence of a real nonzero quadrupole can decrease the aliasing of higher frequencies into the quadrupole, depending on the orientation of the real quadrupole relative to the censored latitudes. At worst, the random component is parallel to the true quadrupole leaving the uncertainty unchanged; at best, the random effect is orthogonal to the true quadrupole, in which case the Pythagorean Theorem bounds the contribution from aliasing. However, the hypothesis that $Q_{rms} = 0$ is consistent with the data. It is not statistically sound to assume that $Q_{rms} > 0$ in order to show that $Q_{rms} > 0$.

Keeping data from a larger range of galactic latitudes, *e.g.*, $|b| > 10°$, decreases the aliasing in the least-squares fit. It would certainly help if the galactic microwave signal could be modelled sufficiently well that data from the entire sky could be used. However, as described in the penultimate paragraph of Section 2, without constraints on CMB fluctuations the uncertainty remains formally infinite. Any simulation of the remaining effect must assume some probabilistic model for CMB heterogeneity. Since there is always a possibility that the probabilistic model is violated, simulations can give only lower bounds on the uncertainty.

Finally, continuing the *COBE* experiment will improve the signal-to-noise ratio consistently, decreasing the random errors of measurement. The contribution of measurement errors to the uncertainty is already less than that of possible aliasing, so constraints are still needed.

*Future Directions.* A variety of techniques are available for finding uncertainties in problems where the unknown is a function obeying *a priori* inequalities (*e.g.* Backus, 1989; Stark, 1992ab; Stark and Hengartner, 1993). Cosmological theories predict essentially zero CMB structure beyond degree $l = 1000$, and other experiments upper-bound CMB heterogeneity on intermediate scales (many are mentioned in Wright et al. 1992; Gaier et al. 1992 give a new bound at an angular scale of 1.2°). These bounds, together with the assumption that the CMB spectrum follows a power law, might permit useful, conservative inferences about the spectral index, even though individual multipoles remain uncertain. One way to do this is to find the range of power laws such that there exists a set of multipole coefficients satisfying goodness-of-fit tests both to the power law and to the data. We plan to apply this approach when the *COBE* data are published (Bunn, Silk, and Stark).

## CONCLUSIONS

While the *COBE* detection of CMB heterogeneity seems solid, estimates of CMB multipole coefficients have such large uncertainties that they are consistent with zero. Strong *a priori* constraints on the nature of CMB heterogeneity are needed to estimate multipole coefficients from the *COBE* data reliably. It is likely that if the spectrum of CMB heterogeneity is assumed to follow a power law, nontrivial confidence intervals on the spectral index and the normalization $Q_{rms-PS}$ could be computed by combining *COBE* data with observational limits from other experiments.

I am extremely grateful to L. Tenorio for introducing me to the CMB problem, for numerous discussions and for computational assistance; and to C.L. Bennett, C. Lineweaver, J. Rice, J. Schuster, J. Silk and G.F. Smoot for helpful discussions. This work was supported by NSF PYI award DMS 8957573.

## REFERENCES

Backus, G. 1989, Geophys. J., 97, 119

Backus, G. & Gilbert, F. 1968, Geophys. J. Roy. Astr. Soc., 16, 169

Bennett, C.L., et al. 1992, ApJ, 396, L7

Gaier, T., et al. 1992, ApJ, 398, L1

Hogan, C.J. 1992, ApJ, 398, L77

Kogut, A., et al., 1992, ApJ, in press

Smoot, G.F., et al. 1992, ApJ, 396, L1

Stark, P.B. 1992a, Geophys. J. Int., 108, 329

Stark, P.B. 1992b, J. Geophys. Res., 97, 14,055

Stark, P.B. & Hengartner, N.W. 1993, J. Geophys. Res., in press

Wickramasinghe, N.C. 1992, Nature, 358, 547

Wright, E.L., et al. 1992, ApJ, 396, L13

P.B. Stark: Department of Statistics, University of California, Berkeley, CA 94720