NOISE CORRELATIONS IN COSMIC MICROWAVE BACKGROUND EXPERIMENTS

SCOTT DODELSON,[1] ARTHUR KOSOWSKY,[1,2] AND STEVEN T. MYERS[3]
Received 1994 March 21; accepted 1994 November 28

ABSTRACT

Many analyses of microwave background experiments neglect the correlation of noise in different frequency or polarization channels. We show that these correlations, should they be present, can lead to severe misinterpretation of an experiment. In particular, correlated noise arising from either electronics or atmosphere may mimic a cosmic signal. We quantify how the likelihood function for a given experiment varies with noise correlation, using both simple analytic models and actual data. For a typical microwave background anisotropy experiment, noise correlations at the level of 1% of the overall noise can seriously *reduce* the significance of a given detection.

Subject headings: cosmic microwave background — methods: data analysis

1. INTRODUCTION

The last few years have witnessed a surge of experiments measuring anisotropies in the cosmic microwave background. The existence of such anisotropies is now firmly established; measurements are becoming plentiful enough to compare different cosmological theories quantitatively. A useful interpretation of a given experiment relies on proper treatment of atmospheric and instrumental noise. This *Letter* focuses on one possible pitfall in analyzing experimental results: noise in different channels of an experiment may be correlated. This correlated noise may mimic a cosmological signal on the sky and significantly alter the interpretation of an anisotropy measurement.

All present anisotropy experiments share several common features. Typically, an experiment measures the deviation of the microwave background temperature from its mean value; this deviation is the temperature anisotropy. Measurements are usually taken at several different frequencies and sometimes at different polarizations for a total of N_c measurements at a given point on the sky. This set of measurements is then repeated in N_p patches on the sky.

The analysis of this type of experiment requires the correlation function, which includes the expected contribution to the signal from cosmological sources, instrumental and atmospheric noise, and foreground sources. In the present work, we ignore the last contribution; foreground sources are considered in detail elsewhere (Brandt et al. 1993; Dodelson & Stebbins 1993; Dodelson & Kosowsky 1994). The correlation function and the data determine the likelihood function (see, e.g., Readhead et al. 1989; Bond et al. 1991), the probability of obtaining the data given a particular theory and the noise parameters. Explicitly, the likelihood function is given by

$$\mathscr{L} = \frac{(2\pi)^{-N/2}}{\sqrt{\det(C)}} \exp\left(-\frac{1}{2} D C^{-1} D\right), \qquad (1)$$

where $N = N_p N_c$ is the total number of data points, C is the $N \times N$ correlation matrix, and D is a N-component vector

[1] NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510-0500.
[2] Departments of Physics and of Astronomy and Astrophysics, Enrico Fermi Institute, University of Chicago, IL 60637-1433.
[3] Division of Physics, Math, and Astronomy, California Institute of Technology, Pasadena, CA 91125.

containing the data. We are interested in small off-diagonal elements in the noise contribution to the correlation matrix.

The actual value of the likelihood function is not significant, but rather its relative value for different potential theories. For the sake of simplicity, we parameterize theories simply by the variance they predict in a given experiment, σ_{th}; then $\mathscr{L}$ is a function of σ_{th}. A maximum in the likelihood function at a nonzero value of σ_{th} marks a detection; the significance of a detection is reflected by the ratio of the value of the likelihood function at its maximum to its value at no theoretical signal. We therefore consider the *likelihood ratio*:

$$\mathscr{R}(\sigma_{\mathrm{th}}) \equiv \frac{\mathscr{L}(\sigma_{\mathrm{th}})}{\mathscr{L}(\sigma_{\mathrm{th}} = 0)}. \qquad (2)$$

In this *Letter* we show analytically how the likelihood ratio changes if small off-diagonal terms are included in the correlation matrix. We conclude that *even small off-diagonal correlations can lead to huge changes in the likelihood ratio, and therefore in the significance of a detection.* This analytical work, in §§ 2 and 3, is useful but perhaps not completely convincing, as it involves certain simplifying assumptions. In § 4 we present the likelihood ratio for the Saskatoon experiment (Wollack et al. 1993), the only measurement of which we are currently aware that reports off-diagonal correlations. The difference between including such correlations in the analysis (as the group properly did) and neglecting them is shown to be dramatic. The Saskatoon experiment, a ground-based apparatus which uses a single HEMT amplifier for each three frequency channels, has large noise correlations, but even for experiments with substantially smaller correlations the difference can still be very important.

2. TWO CHANNEL EXPERIMENT

In this section we illustrate the importance of off-diagonal correlations with a simple example. Consider an experiment which measures the temperature anisotropies in two frequency channels at one point on the sky. In the absence of correlated noise, we need three pieces of information to analyze such an experiment: (i) the data, D, which in this case consists of two numbers, the observed temperature anisotropy in each channel; (ii) the theoretical prediction for the expected rms anisotropy, σ_{th}; and (iii) the expected rms of the noise, σ_n. With two frequency channels, the latter two quantities become 2 × 2

matrices. The correlation matrix is the sum of these two matrices:

$$C_0 = \begin{bmatrix} \sigma_{th}^2 + \sigma_n^2 & \sigma_{th}^2 \\ \sigma_{th}^2 & \sigma_{th}^2 + \sigma_n^2 \end{bmatrix}. \quad (3)$$

The theoretical rms σ_{th}^2 appears in every element, because the expected rms due to the cosmic signal is the same in every channel: if channel 1 measures a given signal d_1, channel 2 is predicted to measure the same value for d_2 in the absence of noise. The theoretically expected signal in each channel is therefore correlated. Any experiment will have diagonal contributions to the noise; for simplicity we assume the same noise rms in each frequency channel. Additional off-diagonal noise components arise whenever the noise sources in different frequency channels are correlated. We parameterize the off-diagonal components by ϵ and write the total correlation matrix as

$$C = \sigma_n^2 \begin{bmatrix} 1 + x & \epsilon + x \\ \epsilon + x & 1 + x \end{bmatrix} \quad (4)$$

where $x \equiv \sigma_{th}^2/\sigma_n^2$. Correlated noise will most likely arise from the atmosphere or from an experiment's electronics.

We can evaluate directly the likelihood function in equation (1) by noting that

$$C^{-1} = \frac{1}{\sigma_n^2(1 - \epsilon)(1 + 2x + \epsilon)} \begin{bmatrix} 1 + x & -\epsilon - x \\ -\epsilon - x & 1 + x \end{bmatrix}. \quad (5)$$

Then writing the two measurements as $D \equiv \sigma_n(\bar{d} + d_1, \bar{d} + d_2)$, where $\bar{d}\sigma_n$ is the mean measurement of the two channels, the likelihood function is

$$\mathcal{L} = \frac{1}{2\pi\sigma_n^2} [(1 - \epsilon)(1 + 2x + \epsilon)]^{-1/2}$$

$$\times \exp\left[-\frac{\bar{d}^2}{1 + 2x + \epsilon} - \frac{d_1^2 + d_2^2}{2(1 - \epsilon)} \right]. \quad (6)$$

A straightforward calculation shows that $\mathcal{L}(x)$ peaks at $x = X$ satisfying

$$2\bar{d}^2 = 1 + 2X + \epsilon \quad (7)$$

if X is greater than zero. This immediately gives the likelihood ratio, defined in equation (2), as

$$\mathcal{R} = \sqrt{\frac{1 + \epsilon}{2\bar{d}^2}} \exp\left(\frac{\bar{d}^2}{1 + \epsilon} - \frac{1}{2} \right). \quad (8)$$

It is instructive to consider a particular limit of equation (8). When the noise in each channel is completely correlated ($\epsilon = 1$), we expect to get less information from this experiment. Instead of two independent channels, we really have only one independent channel. Thus the likelihood ratio in this limit should be the same as for a one channel experiment. A short calculation shows that for a one channel experiment, the likelihood ratio is equal to $(1/\bar{d}) \exp [(2\bar{d}^2 - 1)/2]$, which is indeed the value of $\mathcal{R}$ in equation (8) when $\epsilon = 1$.

We now show that as long as the noise in the two channels is positively correlated, the likelihood ratio decreases. Let us define $\mathcal{R}_0$ to be the likelihood ratio when noise is not correlated ($\epsilon = 0$). Then, equation (8) tells us that

$$\frac{\mathcal{R}}{\mathcal{R}_0} = \sqrt{1 + \epsilon} \exp\left(-\bar{d}^2 \frac{\epsilon}{1 + \epsilon} \right). \quad (9)$$

This ratio is always less than one as long as $0 < \epsilon < 2\bar{d}^2 - 1$. The first inequality ($\epsilon > 0$) holds when the noise is positively correlated; the second holds for any solution of equation (7). Thus *positively correlated noise reduces the likelihood ratio.* This problem is particularly acute in cosmic microwave background experiments, since the signal is also completely correlated in the different channels; thus if the noise is correlated, we expect the statistical significance of detections or upper limits to be weakened.

Even for this simplistic one-patch, two-channel model, the effect of noise correlations can be non-negligible in estimating the significance of a detection. For example, if the signal is twice the noise level ($\bar{d}^2 = 4$), presently a representative signal-to-noise ratio, and if $\epsilon \sim 0.5$ as it is in the Saskatoon experiment, then including the correlation decreases the significance of a detection by almost a factor of three. We will now show that the situation gets worse with multiple channels and patches.

3. GENERALIZATION

It is straightforward to generalize the above discussion to allow for many frequency and/or polarization channels and many spatial patches. For N_c channels, C becomes an $N_c \times N_c$ matrix with components

$$C_{ij} = \sigma_n^2(1 - \epsilon)\left(\delta_{ij} + \frac{x + \epsilon}{1 - \epsilon} \right). \quad (10)$$

Equation (10) idealizes an actual experiment in two ways: (i) the noise in all the channels is assumed equal, so the diagonal components of C are equal; (ii) each pair of channels has equal correlation, so the off-diagonal components of C_1 are all equal. The inverse of C is easily obtained by assuming $C_{ij}^{-1} \propto \delta_{ij} + b$ and imposing the condition $CC^{-1} = 1$ to find the constant b and the normalization; the result is

$$C_{ij}^{-1} = \frac{1}{\sigma_n^2(1 - \epsilon)} \left[\delta_{ij} - \frac{x + \epsilon}{1 + N_c x + (N_c - 1)\epsilon} \right]. \quad (11)$$

The determinant of C is given by

$$\det C = \sigma_n^{2N_c}(1 - \epsilon)^{N_c - 1}[1 + N_c x + (N_c - 1)\epsilon], \quad (12)$$

which can be proven by induction. Combining these two expressions gives the N_c-channel likelihood function:

$$\mathcal{L} = \frac{1}{(2\pi)^{N_c/2}\sigma_n^{N_c}} \{(1 - \epsilon)^{N_c - 1}[1 + N_c x + (N_c - 1)\epsilon]\}^{-1/2}$$

$$\times \exp\left[-\frac{N_c \bar{d}^2/2}{1 + N_c x + (N_c - 1)\epsilon} - \frac{1}{2(1 - \epsilon)} \sum_i d_i^2 \right], \quad (13)$$

which reduces to the previous result for $N_c = 2$. The likelihood function peaks at $x = X$ satisfying

$$N_c \bar{d}^2 = 1 + N_c X + (N_c - 1)\epsilon, \quad (14)$$

and the likelihood ratio result generalizes to

$$\frac{\mathcal{R}}{\mathcal{R}_0} = \sqrt{1 + (N_c - 1)\epsilon} \exp\left\{ -\frac{N_c(N_c - 1)\bar{d}^2\epsilon}{2[1 + (N_c - 1)\epsilon]} \right\}. \quad (15)$$

A further generalization to N_p different patches on the sky can be approximated by a block diagonal correlation matrix, where each block is an identical N_c-channel correlation matrix. This approximation simply raises the right side of equation (15) to the power N_p:

$$\frac{\mathcal{R}}{\mathcal{R}_0} = [1 + (N_c - 1)\epsilon]^{N_p/2} \exp\left\{ -\frac{N_p N_c(N_c - 1)\bar{d}^2\epsilon}{2[1 + (N_c - 1)\epsilon]} \right\}. \quad (16)$$

where $\bar{d}^2$ now represents the mean squared signal-to-noise ratio for all patches. For the noise piece of the correlation matrix, this block diagonal *Ansatz* is normally a good approximation, as noise in different patches is unlikely to be correlated. However taking the full correlation matrix to be block diagonal is only an approximation, since the signal is likely to be correlated from patch to patch unless the patches are far removed from each other compared to the patch size. The error in this approximation does not qualitatively affect our arguments.

For a typical medium angle experiment today, $N_p \simeq 20$ and $N_c \geq 3$. With a signal-to-noise ratio of order 2, the argument of the exponential is typically greater than $200\epsilon/(1 + 2\epsilon)$. If ϵ is of order 0.5, then the significance of a detection decreases by a factor of order $2^{10}e^{-50} \simeq 10^{-19}$ when one accounts for the noise correlations. For $\epsilon \simeq 0.1$, the significance decreases by $\simeq 10^{-7}$, and noise correlated even at the 1% level may reduce the likelihood ratio by nearly a factor of 10. In general, we expect correlations to be important roughly when

$$\epsilon \gtrsim \frac{1}{N_p N_c (N_c - 1)} . \tag{17}$$

As the signal-to-noise ratio increases, the effect of correlations on the likelihood ratio becomes stronger, although the likelihood itself becomes more sharply peaked.

4. SASKATOON EXPERIMENT

In deriving the analytic result in equation (16), we made three assumptions: (1) The noise in each channel and patch was assumed to have the same variance; (2) The noise was assumed to be correlated in the same way between any pair of channels; (3) The signal was assumed to be uncorrelated from one spatial patch to another. In this section we analyze the Saskatoon experiment without making any of these assumptions. The reported error bars (the diagonal variance of the noise) and the reported off-diagonal correlations replace the first two assumptions. We perform the analysis in the context of a standard cold dark matter (CDM) model with a Harrison-Zel'dovich-Peebles initial spectrum, which completely determines the correlations between patches.

The Saskatoon experiment takes measurements at six different channels for each patch: three frequency channels and two polarization channels. We consider the so-called "East" data set, which consists of measurements in 21 separate patches (the "West" data set gives very similar answers). The theory gives the predicted variance not only in a given patch, but also the correlations between different patches. Specifically,

$$\langle d_a d_b \rangle = \sum_{l=2}^{\infty} \frac{2l+1}{4\pi} C_l W_{l,ab} \tag{18}$$

where a, b label different patches; C_l is the prediction of the theory for the lth multiple moment; and $W_{l,ab}$ is the window function of the experiment which depends on the chopping strategy, beam width, and spatial separation of patches a and b. For comparison, we previously assumed $\langle d_a d_b \rangle = \delta_{ab} \sigma_{th}^2$. CDM has only one free parameter, so C_l/C_2 is fixed for all
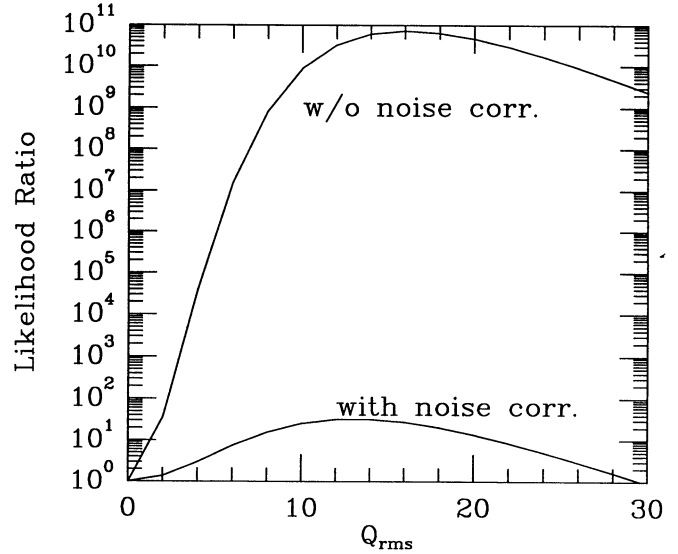

FIG. 1.—Likelihood ratio vs. Q_{rms} for the "East" data of the Saskatoon experiment with an without noise correlations. Q_{rms} is related to C_2 via $Q_{rms}^2 = 5C_2/4\pi$.

$l > 2$. Figure 1 shows the likelihood ratio $\mathscr{R}$ for the Saskatoon East data as a function of $Q_{rms} = (5C_2/4\pi)^{1/2}$.

The two curves in Figure 1 correspond to the likelihood ratio with and without off-diagonal noise correlations. The difference is stunning. A detection which would have been extremely clean $[\mathscr{R}(\bar{C}_2) \sim 10^{11}]$ becomes much less certain $[\mathscr{R}(\bar{C}_2) \sim 30]$ once correlations are accounted for. (Note that eq. [16] actually underestimates the effects of noise correlations in this case: $\sigma_{th}^2/\sigma_n^2 = 0.68$ and $\epsilon \sim \frac{1}{3}$, so that we expect $\mathscr{R}/\mathscr{R}_0 \sim 5 \times 10^{-8}$. One reason for this is that the off-diagonal noise elements were chosen to be equal in the simple model leading up to eq. [16], whereas in any real experiment, certain channels will be more strongly correlated than others.) We emphasize that the Saskatoon experiment *did* include this effect in their analysis; we use this experiment as an illustration because it is the *only* one of which we are aware that has reported the presence (or lack of) noise correlations.

In summary, even relatively small amounts of correlated noise between frequency channels can greatly affect the interpretation of microwave background experiments. The effect is likely to be of particular importance for those experiments which utilize a common part of the electronic signal path for multiple frequency channels (as in the Saskatoon experiment considered here), or for experiments with substantial atmospheric or environmental sensitivity. We hope the arguments presented here will prompt other groups to examine and report noise correlations in their experiments.

REFERENCES

Bond, J. R., Efstathiou, G., Lubin, P. M., & Meinhold, P. R. 1991, Phys. Rev. Lett., 66, 2179
Brandt, W. N., Lawrence, C. R., Readhead, A. C. S., Pakianathan, J. N., & Fiola, T. M. 1993, CalTech-OVRO preprint 12
Dodelson, S., & Kosowsky, A. 1994, astro-ph/9410081
Dodelson, S., & Stebbins, A. 1994, ApJ, 433, 440
Readhead, A. C. S., et al. 1989, ApJ, 346, 566
Wollack, E. J., Jarosik, N. C., Netterfield, C. B., Page, L. A., & Wilkinson, D. 1993, Princeton preprint