Fermi National Accelerator Laboratory

FERMILAB-Pub-94/039-A

astro-ph/9402053

February 1994

Noise Correlations in Cosmic Microwave Background Experiments

Scott Dodelson[1] and Arthur Kosowsky[1,2]

[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
The University of Chicago
Chicago, IL 60637-1433

ABSTRACT

Many analyses of microwave background experiments neglect the correlation of noise in different frequency or polarization channels. We show that these correlations 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 medium angle experiment, noise correlations at the level of 1% of the overall noise can seriously *reduce* the significance of a given detection.

Submitted for publication in *Astrophysical Journal Letters*

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

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.

In brief, all present anisotropy experiments use similar observing strategies. 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 $N_{\rm ch}$ measurements at a given point on the sky. This set of measurements is then repeated in $N_{\rm p}$ 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 and Stebbins, 1993; Dodelson and 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

$$\begin{aligned}\mathcal{L} &= \frac{(2\pi)^{-N/2}}{\sqrt{\det(C)}}\exp\left[-\frac{1}{2}DC^{-1}D\right]\\ &= (2\pi)^{-N/2}\exp\left[-\frac{1}{2}DC^{-1}D-\frac{1}{2}\mathrm{Tr}\ln C\right],\end{aligned}\tag{1}$$

where $N = N_{\rm p}N_{\rm ch}$ is the total number of data points, C is the $N\times N$ correlation matrix, and D is a N-component vector containing the data. We are interested in small off-diagonal elements in the noise contribution to the correlation matrix; it is useful to decompose the correlation matrix into two pieces, $C = C_0 + C_1$, where C_1 is a small perturbation. To first-order in C_1 the likelihood function is given by

$$\mathcal{L} \sim \mathcal{L}_0 \exp\left[\frac{1}{2} D C_0^{-1} C_1 C_0^{-1} D - \frac{1}{2}\mathrm{Tr}(C_0^{-1} C_1)\right]$$
$$\mathcal{L}_0 = \frac{(2\pi)^{-N/2}}{\sqrt{\det C_0}} \exp\left[-\frac{1}{2} D C_0^{-1} D\right] \qquad (2)$$

where $\mathcal{L}_0$ is the likelihood function in the absence of the perturbation. The actual value of the likelihood function is not significant, but rather its relative value for different potential theories. Here we parametrize theories simply by the variance they predict in a given experiment, σ_{th}. Then $\mathcal{L}$ is a function of σ_{th}. A maximum in the likelihood function at a non-zero value of σ_{th} marks a detection; the significance of a detection is given roughly 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*:

$$\mathcal{R}(\sigma_{\mathrm{th}}) \equiv \frac{\mathcal{L}(\sigma_{\mathrm{th}})}{\mathcal{L}(\sigma_{\mathrm{th}} = 0)}. \qquad (3)$$

In this *Letter* we show analytically how the likelihood ratio changes as small off-diagonal correlations are turned on [i.e. as C_1 becomes non-zero]. The conclusion is 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 Sections II and III, is useful but perhaps not completely convincing, as it involves certain simplifying assumptions. In Section IV 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 noise correlations larger than many other current experiments, 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. To analyze such an experiment we need three pieces of information: (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, $\sigma_{\rm th}$; and (iii) the expected rms of the noise, $\sigma_{\rm n}$. With two frequency channels, the latter two quantities become 2×2 matrices. The correlation matrix is the sum of these two matrices:

$$C_0 = \begin{bmatrix} \sigma_{\rm th}^2 + \sigma_{\rm n}^2 & \sigma_{\rm th}^2 \\ \sigma_{\rm th}^2 & \sigma_{\rm th}^2 + \sigma_{\rm n}^2 \end{bmatrix} . \tag{4}$$

The theoretical rms $\sigma_{\rm th}^2$ 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 d_1, channel 2 is predicted to find the same value of 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 parametrize the off-diagonal components by the small parameter ϵ and write

$$C_1 = \begin{bmatrix} 0 & \epsilon\sigma_{\rm n}^2 \\ \epsilon\sigma_{\rm n}^2 & 0 \end{bmatrix} . \tag{5}$$

Correlated noise will most likely arise from the atmosphere or from an experiment's electronics.

We can evaluate the likelihood function in Eq. (2) by noting that

$$C_0^{-1} = \frac{1}{\sigma_{\rm n}^2}\frac{1}{\sigma_{\rm n}^2 + 2\sigma_{\rm th}^2} \begin{bmatrix} \sigma_{\rm n}^2 + \sigma_{\rm th}^2 & -\sigma_{\rm th}^2 \\ -\sigma_{\rm th}^2 & \sigma_{\rm n}^2 + \sigma_{\rm th}^2 \end{bmatrix} . \tag{6}$$

Then writing the two measurements as $D \equiv (d_1, d_2)$, the likelihood function to first order in ϵ is

$$\begin{aligned} \mathcal{L} \simeq & \mathcal{L}_0 \exp\left[\epsilon\left(\frac{d_1 d_2}{\sigma_{\rm n}^2} + \frac{\sigma_{\rm th}^2}{\sigma_{\rm n}^2 + 2\sigma_{\rm th}^2} - \frac{\sigma_{\rm th}^4 + \sigma_{\rm th}^2\sigma_{\rm n}^2}{(\sigma_{\rm n}^2 + 2\sigma_{\rm th}^2)^2}\frac{(d_1+d_2)^2}{\sigma_{\rm n}^2}\right)\right] \\ \mathcal{L}_0 = & \frac{1}{2\pi\sigma_{\rm n}\sqrt{\sigma_{\rm n}^2 + 2\sigma_{\rm th}^2}} \exp\left[\frac{-(d_1^2 + d_2^2)}{2\sigma_{\rm n}^2}\right] \exp\left[\frac{(d_1+d_2)^2}{2}\frac{\sigma_{\rm n}^2}{\sigma_{\rm n}^2 + 2\sigma_{\rm th}^2}\right] . \end{aligned} \tag{7}$$

A straightforward calculation shows that $\mathcal{L}_0$ peaks at

$$\bar{\sigma}_{\rm th}^2 = \frac{1}{4}(d_1 + d_2)^2 - \frac{\sigma_{\rm n}^2}{2} \tag{8}$$

if this quantity is greater than zero. Then substituting this value into Eq. (3) gives the simple result

$$\mathcal{R}(\bar{\sigma}_{\rm th}^2) = \mathcal{R}_0(\bar{\sigma}_{\rm th}^2)\exp(-\epsilon\bar{\sigma}_{\rm th}^2/\sigma_{\rm n}^2), \tag{9}$$

where $\mathcal{R}_0$ is the likelihood ratio when noise is not correlated ($\epsilon = 0$). As long as ϵ is positive, meaning the noise in different channels is positively correlated, the likelihood ratio at $\bar{\sigma}_{\rm th}^2$ is *lowered* by including correlation. Intuitively, correlations in the noise signal can mock those from a cosmic signal; 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 predicted signal is twice the noise level ($\sigma_{\rm th}^2 = 4\sigma_{\rm n}^2$), 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 ten. 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_0 and C_1 become $N_c \times N_c$ matrices:

$$C_0 = \begin{bmatrix} \sigma_{\rm th}^2 + \sigma_{\rm n}^2 & \sigma_{\rm th}^2 & \cdots & \sigma_{\rm th}^2 \\ \sigma_{\rm th}^2 & \sigma_{\rm th}^2 + \sigma_{\rm n}^2 & \cdots & \sigma_{\rm th}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\rm th}^2 & \sigma_{\rm th}^2 & \cdots & \sigma_{\rm th}^2 + \sigma_{\rm n}^2 \end{bmatrix} \tag{10}$$

$$C_1 = \begin{bmatrix} 0 & \epsilon\sigma_{\rm n}^2 & \cdots & \epsilon\sigma_{\rm n}^2 \\ \epsilon\sigma_{\rm n}^2 & 0 & \cdots & \epsilon\sigma_{\rm n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon\sigma_{\rm n}^2 & \epsilon\sigma_{\rm n}^2 & \cdots & 0 \end{bmatrix}. \tag{11}$$

Eqs. (10) and (11) idealize an actual experiment in two ways: (i) the noise in all the channels is assumed equal, so the diagonal components of C_0 are equal; (ii) each pair of channels has equal correlation, so the off-diagonal components of C_1 are all the same. The simplest way to carry out the calculations leading to the likelihood ratio is to write in component form: $(C_0^{-1})_{ij} = -(\sigma^2_{\rm th}/\sigma^2_{\rm n})/(\sigma^2_{\rm n} + N_{\rm ch}\sigma^2_{\rm th}) + \delta_{ij}/\sigma^2_{\rm n}$ and $(C_1)_{ij} = (1-\delta_{ij})\epsilon\sigma^2_{\rm n}$, where δ_{ij} is the Kronecker delta, equal to one when $i = j$ and zero otherwise. Then a straightforward calculation gives the generalizations of Eqs. (8) and (9) as

$$\bar{\sigma}^2_{\rm th} = \left(\frac{1}{N_c}\sum_{i=1}^{N} d_i\right)^2 - \frac{\sigma^2_{\rm n}}{N_{\rm ch}} \tag{12}$$

and

$$\mathcal{R}(\bar{\sigma}^2_{\rm th}) = \mathcal{R}_0(\bar{\sigma}^2_{\rm th})\exp\left[-\frac{N_{\rm ch}(N_{\rm ch}-1)\bar{\sigma}^2_{\rm th}\epsilon}{2\sigma^2_{\rm n}}\right]. \tag{13}$$

Both of these expressions reduce to the two-channel result in Eq. (8) and (9). The effect of noise correlation is enhanced by the factor $N_{\rm ch}(N_{\rm ch}-1)/2$.

A further generalization to many different patches on the sky can be approximated by a block diagonal correlation matrix:

$$\mathcal{C}_0 = \begin{bmatrix} C_0 & 0 & \dots & 0 \\ 0 & C_0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_0 \end{bmatrix} \tag{14}$$

and

$$\mathcal{C}_1 = \begin{bmatrix} C_1 & 0 & \dots & 0 \\ 0 & C_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_1 \end{bmatrix}. \tag{15}$$

Here C_0 and C_1 are the $N_{\rm ch} \times N_{\rm ch}$ matrices given in Eqs. (10) and (11); each is repeated $N_{\rm p}$ times to compose the $N \times N$ matrices $\mathcal{C}_0$ and $\mathcal{C}_1$. For the noise correlation matrix $\mathcal{C}_1$ this block diagonal form is probably a good approximation, as noise in different patches is unlikely to be correlated. However taking $\mathcal{C}_0$ to be block diagonal is only an approximation, since the signal is likely to be correlated from patch to patch unless the patches are very far removed from each other. This approximation does not qualitatively affect our arguments.

The corresponding forms of Eqs. (8) and (9) now become

$$\bar{\sigma}_{\rm th}^2 = \frac{1}{N_{\rm p} N_{\rm ch}^2} \sum_{a=1}^{N} \left(\sum_{i=1}^{N} d_{i,a} \right)^2 - \frac{\sigma_{\rm n}^2}{N_{\rm ch}} \tag{16}$$

and

$$\mathcal{R}(\bar{\sigma}_{\rm th}^2) = \mathcal{R}_0(\bar{\sigma}_{\rm th}^2) \exp\left[-N_{\rm p} \frac{N_{\rm ch}(N_{\rm ch}-1)\bar{\sigma}_{\rm th}^2 \epsilon}{2\sigma_{\rm n}^2} \right]. \tag{17}$$

Eqs. (16) and (17) are our final analytic results. For a typical medium angle experiment today, $N_{\rm p} \simeq 20$ and $N_{\rm ch} \simeq 3-4$. Even with a low signal to noise ratio of order one, the argument of the exponential is then of order 100ϵ. If ϵ is of order 0.1, then the significance of a detection decreases by a factor of order $e^{10} \simeq 10^4$ when one accounts for the noise correlations. In general, for low signal-to-noise ratios, we expect correlations to be important if

$$\epsilon \gtrsim \frac{1}{N_{\rm p} N_{\rm ch}(N_{\rm ch}-1)}; \tag{18}$$

as the signal-to-noise ratio increases, the effect of correlations on the likelihood ratio becomes stronger.

4. Saskatoon Experiment

In deriving the analytic result in Eq. (17), 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 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,

$$\langle d_a d_b \rangle = \sum_{l=2}^{\infty} \frac{2l+1}{4\pi} C_l W_{l,ab} \tag{19}$$

where a, b label different patches; C_l is the prediction of the theory for the l^{th} mulitple 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 $\langle d_a d_b \rangle = \delta_{ab}\sigma_{\mathrm{th}}^2$. CDM has only one free parameter, so C_l/C_2 is fixed for all $l > 2$. Figure 1 shows the likelihood ratio $\mathcal{R}$ for the Saskatoon East data as a function of $Q_{\mathrm{rms}} = \sqrt{5C_2/4\pi}$.

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 $[\mathcal{R}(\bar{C}_2) \sim 10^{11}]$ becomes much less certain $[\mathcal{R}(\bar{C}_2) \sim 30]$ once correlations are accounted for. (Note that Eq. (17) overestimates the effects of noise correlations in this case: $\sigma_{\mathrm{th}}^2/\sigma_{\mathrm{n}}^2 = .68$ and $\epsilon \sim 1/3$, so that the argument of the exponential is ~ 70. This is because the off-diagonal elements in C_1 were chosen to be equal in the simple model leading up to Eq. (17) , whereas in any real experiment, certain channels will be more strongly correlated than others.) We emphasize that the observers *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 noise correlations. We hope the arguments presented here will prompt other groups to examine and report correlations in their experiments.

We are especially grateful to Lyman Page and Stephan Meyer for their patient explanation of the Saskatoon experiment, and to them and Michael Turner for other useful discussions. This work was supported in part by the DOE (at Chicago and Fermilab) and by NASA through grant No. NAGW-2381 (at Fermilab). AK is supported in part by the NASA Graduate Student Researchers Program.

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., and Fiola, T.M. 1993, CalTech-OVRO preprint # 12

Dodelson, S. and Kosowsky, A., 1994, work in progress.

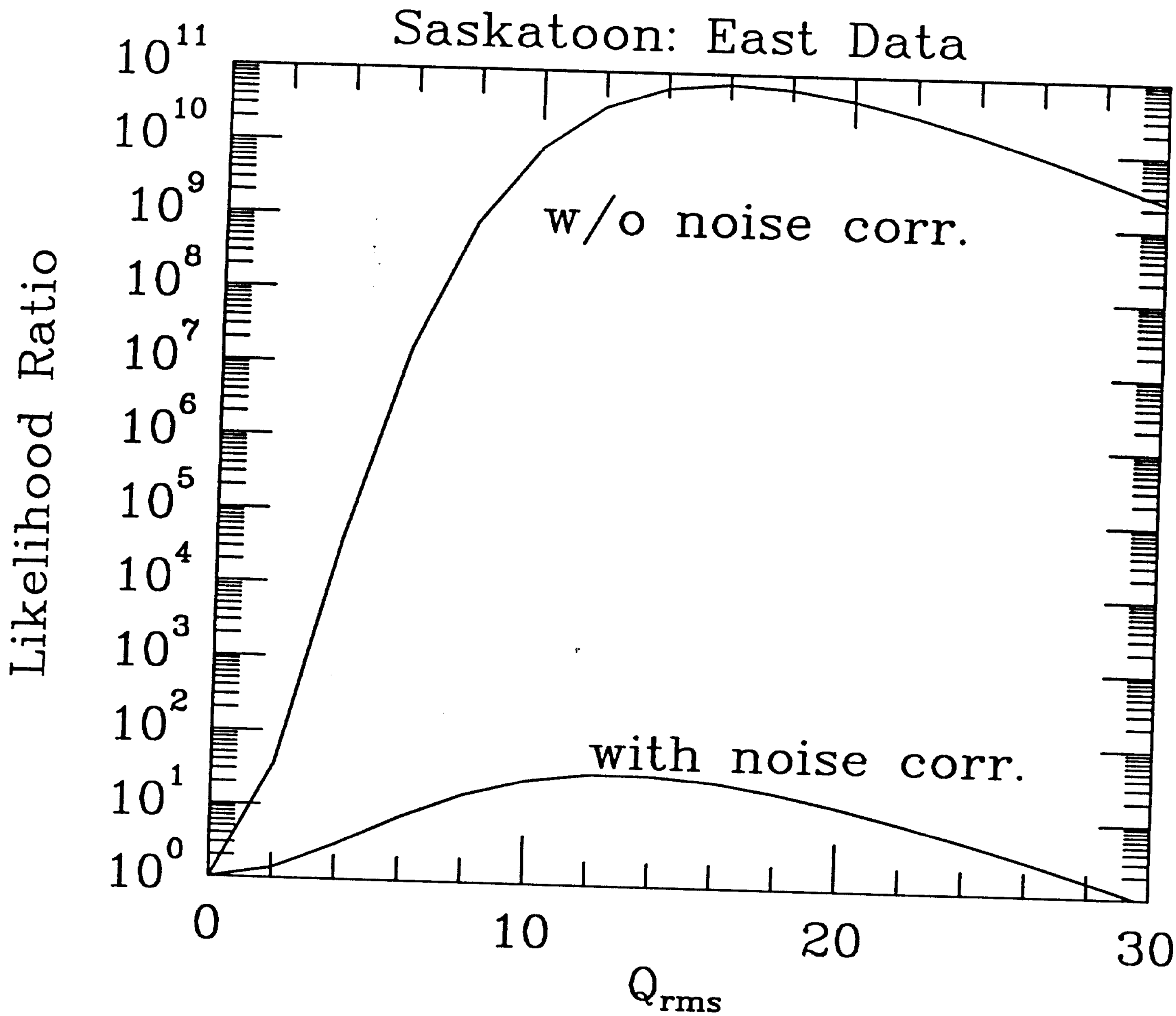
Dodelson, S. & Stebbins, A., 1993 Fermilab preprint, astro-ph/9310048.

Readhead, A. C. S. *et al.*, 1989, *Ap. J.*, **346**, 566.

Wollack, E.J., Jarosik, N.C., Netterfield, C.B., Page, L.A., & Wilkinson, D., 1993, Princeton preprint.

Figure Captions

Figure 1. The Likelihood Ratio vs. $Q_{\rm rms}$ for the East data of the Saskatoon experiment with and without noise correlations. $Q_{\rm rms}$ is related to C_2 via $Q^2_{\rm rms} = 5C_2/4\pi$.


Saskatoon: East Data
Likelihood Ratio
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}
10^{6}
10^{7}
10^{8}
10^{9}
10^{10}
10^{11}
w/o noise corr.
with noise corr.
0
10
20
30
Q_{rms}