Comparing cosmic microwave background datasets

L. Knox,[1] J. R. Bond,[1] A. H. Jaffe,[2] M. Segal,[1] and D. Charbonneau[1]

[1]*Canadian Institute for Theoretical Astrophysics, Toronto, Ontario, Canada M5S 3H8*

[2]*Center for Particle Astrophysics, 301 LeConte Hall, University of California, Berkeley, California 94720*

To extract reliable cosmic parameters from cosmic microwave background datasets, it is essential to show that the data are not contaminated by residual non-cosmological signals. We describe general statistical approaches to this problem, with an emphasis on the case in which there are two datasets that can be checked for consistency. A first visual step is the Wiener filter mapping from one set of data onto the pixel basis of another. For more quantitative analyses, we develop and apply both Bayesian and frequentist techniques. We define the ''contamination parameter'' and advocate the calculation of its probability distribution as a means of examining the consistency of two datasets. The closely related ''probability enhancement factor'' is shown to be a useful statistic for comparison; it is significantly better than a number of χ^2 quantities we consider. Our methods can be used internally (between different subsets of a dataset) or externally (between different experiments), for observing regions that completely overlap, partially overlap or overlap not at all, and for observing strategies that differ greatly. We apply the methods to check the consistency (internal and external) of the MSAM92, MSAM94 and Saskatoon Ring datasets. From comparing the two MSAM datasets, we find that (given a particular model of the contaminant) the most probable level of contamination is 12%, with no contamination only 1.05 times less probable, 50% contamination about 8 times less probable and 100% contamination strongly ruled out at over 2×10^5 times less probable. From comparing the 1992 MSAM flight with the Saskatoon data, we find the most probable level of contamination to be 50%, with no contamination only 1.6 times less probable and 100% contamination 13 times less probable. Our methods can also be used to calibrate one experiment off of another. To achieve the best agreement between the Saskatoon and MSAM92 data, we find that the MSAM92 data should be multiplied by (or Saskatoon data divided by) $1.06^{+0.22}_{-0.26}$. [S0556-2821(98)05320-X]

I. INTRODUCTION

The cosmic microwave background (CMB) is blackbody radiation with a mean temperature of 2.728 ± 0.004 K with 95% confidence [1]. This mean is modulated by a dipole due to our peculiar motion with respect to the radiation field. If one removes the dipole, the temperature is uniform in every direction to ± 100 μK. Precision measurement of these tiny deviations from isotropy can tell us much about the Universe [2].

Unfortunately, precision measurement of 100 μK fluctuations is not an easy task. Even given sufficient detector sensitivity and observing time, one still has to contend with many possible contaminants such as side lobe pickup of the 300° Kelvin Earth and atmospheric noise (even from high-altitude balloons). In addition, there can be contamination of CMB observations by astrophysical foregrounds.

Despite these difficulties, there is good reason to believe that, at least for some experiments, the signals observed from sub-orbital platforms are not dominated by contaminants. One of the best reasons for believing this comes from the comparisons that have been done—between FIRS and Differential Microwave Radiometer (DMR) [3], Tenerife and DMR [4], MSAM92 and Saskatoon [5], MSAM92 and MSAM94 [7], and two years of Python data [6]. Especially for the case when data being compared are from two different instruments, almost the only thing their acquisitions have in common is that they were observing the same piece of sky—each dataset has entirely different sources of systematic error.

In addition to confirming the astrophysical origin of the estimated signal, comparison can greatly improve the ability to detect foreground contamination. Perhaps the best evidence for the thermal nature of anisotropy comes from the comparison between the MSAM92 and Saskatoon datasets. Together, these observations span a frequency range from 36 GHz to greater than 170 GHz. In [5] it was found that the spectral index β ($\delta T \propto (\nu/\nu_0)^\beta$) is constrained to be $\beta = -0.1 \pm 0.2$. For CMB, free-free and dust over this frequency range, we expect $\beta = 0, -1.45$ and 2.25 respectively. The authors conclude that the signals (in the region of overlap) are not dominated by contamination from known astrophysical foregrounds, but are, rather, primarily CMB.

We should not let this apparent success fool us into thinking that going to the next level of precision will be easy. There is a big difference in the level of toleration of contaminants when the goal switches from detection to precision measurement. It is likely that there will be significant levels of contamination (from the atmosphere, side lobes, and foregrounds) in future missions. It may be difficult to demonstrate convincingly that contamination is low without comparison.

Given the importance of comparison, we feel it is worth improving upon the methods used previously. Past treatments have had to ignore much relevant data, and make uncontrolled approximations. This is due to the fact that generally the two datasets being compared were obtained from

instruments observing the sky in different ways. The beam patterns and differencing schemes may differ as in the case of the MSAM92/Saskatoon comparison. In [5] one of the MSAM92 differencing schemes was approximately recreated in software in order to do the comparison. However, no use of software could change the fact that the MSAM92 and Saskatoon beam patterns, although they have fairly similar full-widths at half-maximum, differ significantly in shape. Even when the differencing schemes and beam patterns are the same, there can still be barriers to a direct comparison. The two MSAM flights (one in 1992 called MSAM92 and one in 1994 called MSAM94) took data with essentially the same beam pattern and applied the same differencing, but in this case the direct comparison is frustrated by the fact that the pixels do not all line up exactly. Therefore in [7], pixels within half a beam width of each other were approximated as being at the same point, and those pixels with no partner from the other dataset within this distance were ignored. Half of the data were lost this way.

Here we develop methods of comparing datasets that do not require any information to be thrown away. Differences in demodulation schemes, and effects due to non-overlapping pixels are automatically taken into account. The inevitable price we pay for this is model-dependence. However, we generally expect the model-dependence to be small and indeed find it to be so in the case studies shown here.

An extremely useful tool for visual comparison is the Wiener filter. Roughly speaking, it allows us to interpolate the results from one experiment onto the expected results for another experiment that has observed the sky differently. After some notational preliminaries in Sec. II, in Sec. III, we introduce the Wiener filter in the context of the probability distribution of the signal, given the data. Also in this section, we describe the datasets and apply the Wiener filter to them.

When comparing datasets, we are testing the consistency of our model of the datasets. We emphasize that meaningful model consistency testing demands the existence of other models with which to compare. Therefore we extend our model of the data to include a possible contaminant and calculate the probability distribution of its amplitude, given the data. We find a more limited extension of the model space to also be useful, in which we only consider one alternative to no contamination: complete contamination. We define the ''probability enhancement factor'' as the logarithm of the ratio of the probability of no contamination to the probability of complete contamination. This Bayesian approach to comparison is described and applied in Sec. IV.

In Sec. V we discuss and apply frequentist techniques such as χ^2 tests. The probability enhancement factor can also be used as the basis for a frequentist test—and it is in fact the well-known likelihood ratio test. We demonstrate that the probability enhancement factor has more discriminatory power than any of the other tests considered.

After a further look at the data with the probability enhancement factor in Sec. VI, we discuss the fixing of relative calibration in Sec. VII and possible contamination due to dust in Sec. VIII. Finally we summarize our results in Sec. IX.

II. PRELIMINARIES

Before moving on to a discussion of the various statistics to be used in comparing datasets, we give some review which will serve to define our notation, following Ref. [8].

In general, CMB observations are reduced to a set of binned observations of the sky, or pixels, Δ_i, $i=1...N$ together with a noise covariance matrix, $C_{nii'}$. We model the observations as contributions from signal and noise,

$$\Delta_i = s_i + n_i. \tag{2.1}$$

We assume that the signal and noise are independent with zero mean, with correlation matrices given by

$$C_{Tii'} = \langle s_i s_{i'} \rangle; \quad C_{nii'} = \langle n_i n_{i'} \rangle \tag{2.2}$$

so

$$\langle \Delta_i \Delta_{i'} \rangle = C_{Tii'} + C_{nii'} \tag{2.3}$$

where $\langle ... \rangle$ indicate an ensemble average. With the further assumption that the data are Gaussian, these two-point functions are all that is necessary for a complete statistical description of the data.

One important complication to the above description comes from the existence of constraints. Often the data, Δ_i, are susceptible to some large source of noise, or a not-well-understood source of noise that contaminates only one mode of the data. For example, there may be an unknown offset in the data. In this case, the average is usually subtracted from Δ_i. Similarly, the monopole and dipole are explicitly subtracted from the all-sky COBE/DMR data, because the monopole is not determined by the data and the dipole is local in origin. In general, placing any constraint on the data or some subset thereof, such as insisting that its average be zero, results in additional correlations in Δ_i. We take this into account by adding these additional correlations, C_C, to the noise matrix to create a ''generalized noise matrix,'' C_N, where $C_N = C_n + C_C$. In the limit that the amplitude of C_C gets large, this is equivalent to projecting out those modes which are now unconstrained by the data, but we find this scheme simpler to implement numerically. Thus in the text below we always write the noise matrix as C_N instead of C_n. The details of this procedure for handling the effect of constraints are explained in [8].

Due to finite angular resolution and switching strategies designed to minimize contributions from spurious signals (such as from the atmosphere), the signal is generally not simply the temperature of the sky in some direction, $T(\hat{x})$, but a linear combination of temperatures:

$$s_i = \int d^2\hat{x} H_i(\hat{x}) T(\hat{x}) \tag{2.4}$$

where $H_i(\hat{x})$ is sometimes called the ''beam map,'' ''antenna pattern'' or ''synthesis vector.'' If we discretize the temperature on the sky, then we can write the beam map in matrix form, $s_i = \Sigma_n H_{in} T_n$.

The temperature on the sky, like any scalar field on a sphere, can be decomposed into spherical harmonics

$$T(\theta,\phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta,\phi). \quad (2.5)$$

If the anisotropy is *statistically* isotropic, i.e., there are no special directions in the mean, then the variance of the multipole moments, $a_{\ell m}$, is independent of m and we can write:

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell \ell'} \delta_{m m'}. \quad (2.6)$$

For theories with statistically isotropic Gaussian initial conditions, the angular power spectrum, C_ℓ, is the entire statistical content of the theory in the sense that any possible predictions of the theory for the temperature of the microwave sky can be derived from it.[1]

The theoretical covariance matrix, $C_{Tii'}$, is related to the angular power spectrum by

$$C_{Tii'} = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell W_{ii'}(\ell), \quad (2.7)$$

where

$$W_{ii'}(\ell) = \sum_{nn'} H_{in} H_{i'n'} P_\ell(\cos\theta_{nn'}) \quad (2.8)$$

is called the window function of the observations and $\theta_{nn'}$ is the angular separation between the points on the sphere labeled by n and n'.

Within the context of a model, the C_ℓ depend on some parameters, a_p, $p=1...N_p$ which could be the Hubble constant, baryon density, redshift of reionization, etc. The theoretical covariance matrix will depend on these parameters through its dependence on C_ℓ. We can now write down the probability distribution for the data, given the model parameters, a_p:

$$P(\Delta|C_T(a_p)I) = \frac{1}{(2\pi)^{N/2}|C_T(a_p)+C_N|^{1/2}}$$

$$\times \exp\left(-\frac{1}{2}\Delta^T(C_T(a_p)+C_N)^{-1}\Delta\right).$$

$$(2.9)$$

The I here stands generically for information—in this case the information that the noise is Gaussian-distributed with zero mean and variance C_N.

---

[1] Non-linear evolution will produce non-Gaussianity from Gaussian initial conditions, but this is quite sub-dominant for $\ell \lesssim 1000$.

III. WIENER FILTERS

A. Derivation

Bayes' theorem [9]

$$P(s|\Delta I) = P(s|I)P(\Delta|sI)/P(\Delta|I) \quad (3.1)$$

follows from elementary rules of probability. If we take $P(s|I)$ to be a Gaussian distribution with zero mean and covariance C_T and $P(\Delta|sI)$ to be a Gaussian with mean s and variance C_N, then with a little algebra it follows that the probability distribution for the signal, given the data, C_T and C_N, is

$$P(s|\Delta,C_T,C_N) = \frac{\exp[-(1/2)(s-w\Delta)^\dagger M^{-1}(s-w\Delta)]}{[(2\pi)^{N/2}|M|^{1/2}]},$$

$$(3.2)$$

where $M \equiv \langle(s-w\Delta)(s-w\Delta)^\dagger\rangle = C_T - wC_T$ and

$$w \equiv C_T(C_T+C_N)^{-1} \quad (3.3)$$

is the *Wiener filter* [10]. As one can immediately see from Eq. (3.2), the most probable value of the signal is given by $w\Delta$. As with all Gaussian distributions, this most probable value is also the mean: $\bar{s} \equiv \int s P(s|\Delta,C_T,C_N)ds = w\Delta$.

Thus the Wiener filter operating on the data provides us with the most probable estimate of the underlying signal. Of course, this is the most probable signal only once we assume a power spectrum, C_l, which is used to calculate C_T. Fortunately, this model dependence is quite weak: the Wiener filter provides a robust estimate of the underlying signal provided theories are not chosen which are clearly incompatible with the data.

The Wiener filter can be very helpful for visualizing the underlying signal. For example, often the data are oversampled; that is, there are closely spaced data points with plenty of scatter and large error bars. In a sense, the Wiener filter knows that the high spatial frequency scatter is due to noise and not signal and performs a smoothing of the data—an interpolation controlled by the different statistical properties of the noise and signal.

One can also use the dataset to calculate the most probable signal in some other dataset;[2] let us call the two datasets Δ_1 and Δ_2, where the subscripts refer here to the entire appropriate data vector, not the single element at a particular pixel. Before getting to $P(s_2|\Delta_1)$, we describe some notation for joint datasets. We represent the total data vector as

$$\Delta \equiv \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}. \quad (3.4)$$

This vector will have a total covariance matrix

---

[2] In Ref. [18] the Wiener filter was used to calculate the most probable signal in the Tenerife data, given the Cosmic Background Explorer (COBE) DMR data.

$$\langle \Delta \Delta^{\dagger} \rangle = \begin{pmatrix} \langle \Delta_1 \Delta_1^{\dagger} \rangle & \langle \Delta_1 \Delta_2^{\dagger} \rangle \\ \langle \Delta_2 \Delta_1^{\dagger} \rangle & \langle \Delta_2 \Delta_2^{\dagger} \rangle \end{pmatrix}$$

$$= \begin{pmatrix} C_{T11} + C_{N11} & C_{T12} \\ C_{T21} & C_{T22} + C_{N22} \end{pmatrix} \qquad (3.5)$$

where C_{Tij} represents the theoretical covariance between the pixels of experiments i and j, and $C_{Tij} = C_{Tji}^{\dagger}$. We will also define $C_{ij} = C_{Tij} + C_{Nij}$. We assume that the experiments have no common noise sources (as would be the case for instrument noise and atmospheric noise) and thus $C_{N12} = 0$.

With this notation established we can now write

$$P(s_2 | \Delta_1, C_T, C_N)$$

$$= \frac{\exp[-(1/2)(s_2 - w_{21}\Delta_1)^{\dagger} M^{-1}(s_2 - w_{21}\Delta_1)]}{[(2\pi)^N |M|]^{1/2}}$$

$$(3.6)$$

where $M = C_{T22} - w_{21}C_{T12}$,

$$w_{21} \equiv C_{T21}(C_{T11} + C_{N11})^{-1} \qquad (3.7)$$

and we refer to the Wiener-filtering of dataset one "onto" dataset two.

Thus Wiener-filtering provides us with an excellent tool for visual comparison of datasets. Even if each dataset is expressed in different generalized pixels, since we can Wiener filter one onto the other, we can compare the signal predictions in the same space. We will see applications of this following the next section, which describes the MSAM and Saskatoon datasets.

The Wiener filter can also be derived without reference to anything other than the two-point correlation function of the signal and noise. Assume we want w to be such that the variance $\langle (s - w\Delta)(s - w\Delta)^{\dagger} \rangle$ is minimal. Differentiating with respect to w_{ij}, setting to zero and solving for w_{ij} results in $w = C_T(C_T + C_N)^{-1}$. Thus the minimum-variance estimate of the signal does not depend on the Gaussianity of the signal and noise distributions. Although, of course, the uncertainty in the estimate does [10].

One final expression we will need below is the probability distribution for the data itself, Δ_2 (as opposed to the *signal* in the second dataset) given Δ_1 and relevant matrices. It is the same as the above after changing s_2 to Δ_2 and M to $M + C_{N22}$.

B. Applications

For Gaussian signal and noise, the Wiener filter provides the maximum-likelihood reconstruction of the signal; it is also optimal in the minimum-variance sense discussed above. One can construct a Wiener filter from the pixelized data space onto the same space or from the pixelized data space to any other linear combination of map pixels—such as the map pixels themselves. Wiener filter maps have been made for the SK dataset [11] and the COBE/DMR dataset [12]. Map-making though is not the most useful means for
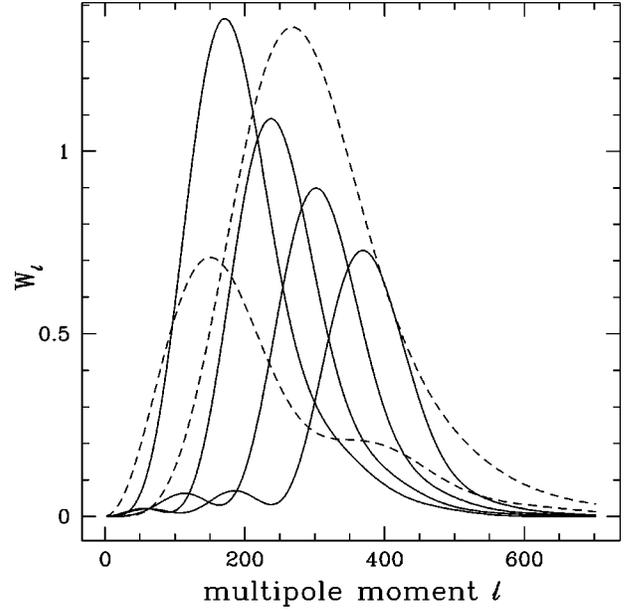
FIG. 1. The diagonal elements of the window function matrix W_{lij} for the four SK95 ring antenna patterns (solid) and the two MSAM antenna patterns (dashed) which are the same for MSAM92 and MSAM94. These show how the power spectrum contributes to the variance of the data [see Eq. (2.7)].

comparing observations that are not themselves maps, and it is not suggested by the statistical techniques we discussed earlier. Here we Wiener filter onto the experimental pixel space itself.

1. Description of the datasets

Before jumping into the applications to the Saskatoon and MSAM datasets, we must describe them. They have considerable spatial overlap and similar angular resolutions. Otherwise, however, the two datasets are very different and a comparison provides a strong check on systematic errors.

MSAM is a balloon-borne bolometric instrument with approximately half-degree (fwhm) resolution in 4 frequency bands centered at 170, 280, 500 and 680 GHz [13]. The data, at each frequency, are binned into pixels on the sky with two different antenna patterns, H, referred to as 2-beam and 3-beam or single-difference and double-difference (see corresponding window functions in Fig. 1). Simultaneously, long time-scale drifts are removed which has the effect of introducing off-diagonal noise correlations. From this multi-frequency data, a fit is made to temperature fluctuations about a 2.73 K black-body component and the optical depth of a dust component. The dust is assumed to have a temperature of 20 K and emissivity that varies with frequency to the 1.5 power.

The MSAM instrument flew in 1992 [14], 1994 [15] and 1995 [16]. In each year a narrow strip of sky with nearly constant declination was observed. The purpose of the 1994 flight was to confirm the results from the 1992 flight and so each targeted the same strip of sky at $\delta = 82°$ (see Fig. 1). Note that, due to, for example, imperfect pointing control, the two flights have slightly different sky coverage. The final
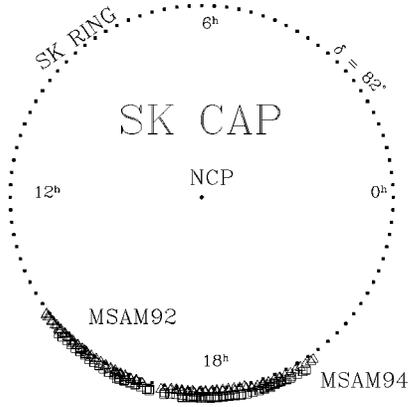
FIG. 2. Observation locations. The SK95 ring data covered the entire circle of radius 8 degrees around the NCP. The centers of the MSAM92 (MSAM94) pixels are indicated with triangles (squares).

flight in 1995 observed near declination $\delta = 80.5°$, chosen to be sufficiently far away from the first two flights for the signal correlations to be negligible. Therefore we do not consider the 1995 flight any further in this paper.

The SK data are reported as complicated chopping patterns (i.e., beam patterns, H, above) in a circle of radius about 8° around the North Celestial Pole. The data were taken over 1993–1995. Here we only use the 1995 data which were taken with angular resolution 0.5° FWHM at approximately 40 GHz. More details can be found in [5].

The bulk of the data were in the ''cap'' configuration: constant-elevation scans tracing out curved rays from the pole, which were then binned in RA and subjected to various sinusoidal demodulation templates in software. Some of the 1995 data (0.5° beam), however, were taken in the ''ring'' configuration, which isolated the data taken at $\delta = 82°$, put into 96 RA bins, and then subjected to 3, 4, 5 and 6 point sinusoidal demodulations, this time along lines of constant declination. The ring data window functions are in Fig. 1. The region of overlap of the SK95 ring data with the two MSAM datasets can be seen in Fig. 2. Throughout when we refer to the SK95 data, we mean only the ring data.

The calibration of the SK dataset was performed by comparing with the supernova remnant Casseiopia A. This star's 30–40 GHz flux itself is poorly determined; hence, the original SK dataset was reported with a 14% calibration error. More recently, Leitch [17] in turn used the brightness of Jupiter (known to 3% in this band) to determine the absolute flux of Cas A; this has resulted in a 5% increase in the temperature of the SK data (and errors), with a reduced calibration error of 7% (the flux of Cas A itself is now determined to $\sim 5\%$, but there are additional sources of calibration error [18]). Except for Sec. VII, in the following we do *not* include the effects of calibration uncertainty.

2. Wiener-filtering MSAM92

An example of Wiener filtering with Eq. (3.3) is shown in Fig. 3. The data points are the values of the pixelized data, located horizontally according to the right ascension of the center of the pixel. The dependence of the pixels on declination and twist has been suppressed. The error bars are from
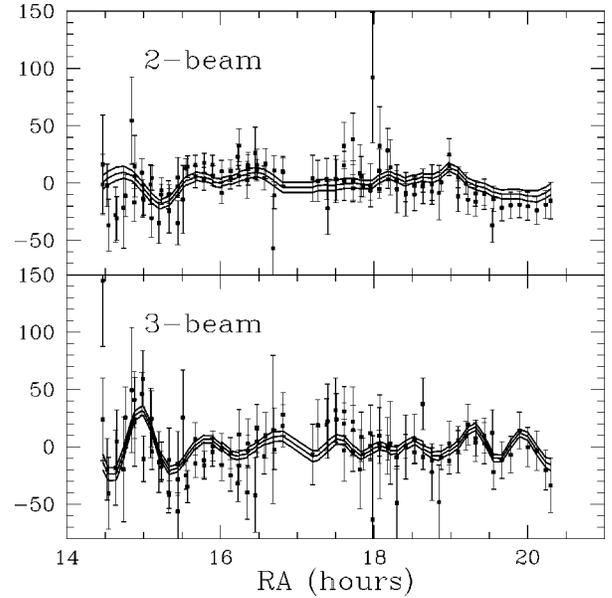
FIG. 3. An example of Wiener filtering. The points with error bars are the MSAM92 pixelized data. Two-beam in top panel, 3-beam in bottom panel. The three curves in each panel are the Wiener-filtered data bounded by $\pm$ one standard deviation.

the diagonal part of the (non-diagonal) noise covariance matrix. The central curve is the Wiener-filtered data and the bounding curves indicate the 68% confidence region for the signal. Because of the difference between the noise covariance matrix and the signal matrix, the Wiener filter essentially assumes that the high frequency behavior is noise and therefore smooths out the data. This smoothing is complicated by the off-diagonal noise correlations which explains some apparent disagreements between the data and the Wiener-filtered data. For example, around 20 hours in the top panel, the Wiener-filtered data are consistently above a number of the data points.

The Wiener filter is model-dependent—one must know (or assume) covariance matrices for the noise and signal. Presumably, the noise covariance matrix is well-known and so the model-dependence resides in the choice of angular power spectrum. Of course, we can gain some knowledge of the angular power spectrum by performing a likelihood analysis of the data. The Wiener filter is generally quite robust to changes in the angular power spectrum that are smaller than those that significantly alter the likelihood—even large changes usually have very little effect. We demonstrate this robustness here with Fig. 4 which shows the Wiener-filtered data for a standard CDM spectral shape and also for a flat spectrum ($\mathcal{C}_\ell$ = constant).

C. Wiener-filtering MSAM94 onto MSAM92

Besides Wiener filtering the data onto its own pixel space, we can Wiener filter it onto another pixel space [Eq. (3.7)]. This provides an excellent visual tool for checking compatibility of results. We show this first for the Wiener filtering of MSAM94 onto MSAM92, together with MSAM92 onto MSAM92 from the previous subsection. Notice that in Fig. 5
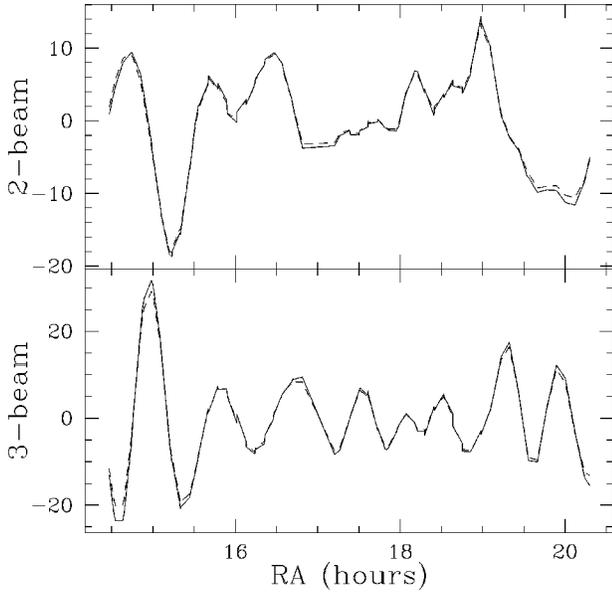
FIG. 4. Wiener filter model-dependence for MSAM92. The standard CDM (flat) spectrum was assumed for the solid (dashed) curve.

the 68% confidence regions mostly overlap each other. One can see the MSAM94 region get wider at either extreme in RA. This is because the MSAM94 pixels have a slightly shorter RA extent than the MSAM92 pixels (14.9^h to 20.1^h compared to 14.5^h to 20.3^h).

One can see in the figure that many features are seen by both datasets; they agree quite well. The most significant differences between the two estimates of the signal are in the region of 15.5 hours for the 2-beam signal and 14.5 hours for
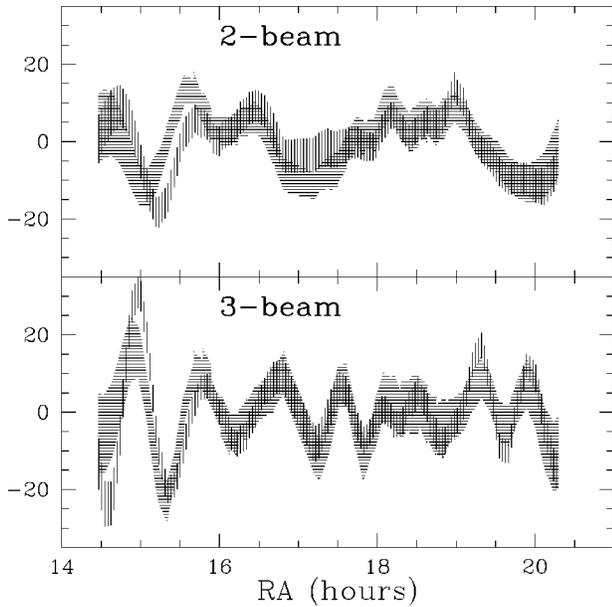
FIG. 5. Wiener filters onto MSAM92 pixels for MSAM92 data (vertical lines) and MSAM94 data (horizontal lines). The curves are realizations consistent with the MSAM94 data. Two-beam in top panel, three-beam in bottom panel.
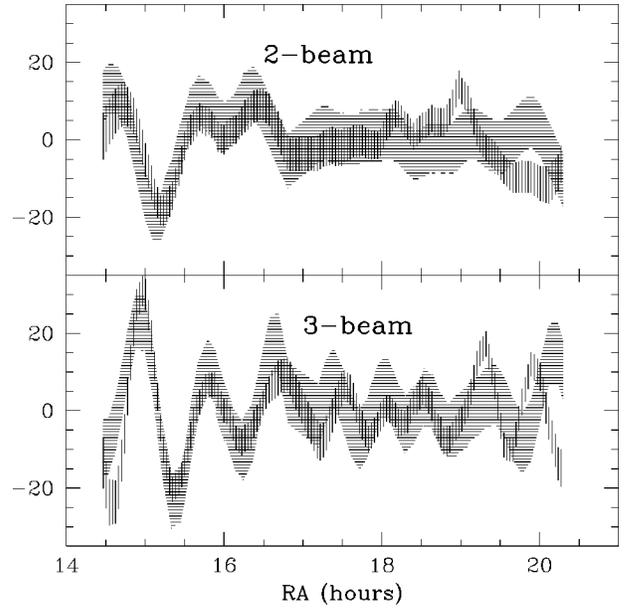
FIG. 6. The MSAM92 (vertical lines) and MSAM95 data (horizontal lines) Wiener-filtered onto the MSAM92 pixels.

the 3-beam signal. We will discuss these slight anomalies later.

D. SK95 onto MSAM92 and MSAM92 onto SK95

Figure 6 shows the same thing as Fig. 5 except that MSAM94 has been replaced with SK95. Once again, the first impression is of general agreement, although the discrepancies here (at large RA) appear to be more significant than those seen in the MSAM92/MSAM94 comparison.

We can also filter the MSAM92 data onto the four SK95 (ring) templates, as shown in Fig. 7. We have chosen the range of this plot to extend in RA further than the MSAM92
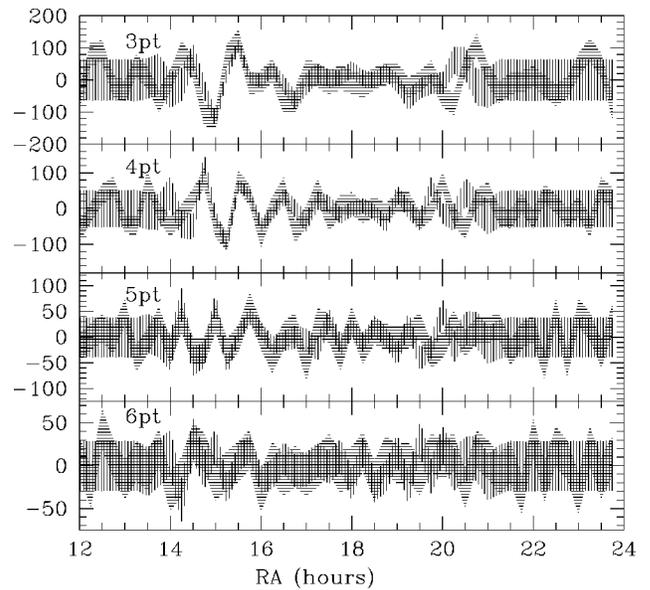
FIG. 7. Wiener filters onto SK95 pixels for MSAM92 data (vertical lines) and SK95 data (horizontal lines).

coverage. This allows one to see how the constraint behaves outside of the region of MSAM92's influence. Notice that the errors do not become infinite. This is because of the prior information that went into the estimate of the probability distribution, i.e., the assumed power spectrum. Also note that the data have some influence slightly beyond the limit of the sky coverage. The dominant reason for this is the spatial extent of the antenna patterns. In addition, the intrinsic correlations (assumed in the prior) extend the influence to slightly beyond where the antenna response is zero.

With two dimensional Wiener filter maps (as with any two dimensional map), it is difficult to plot both the map and a confidence region expressing the level of uncertainty as we have done here for essentially one dimensional data. In 2D it is therefore often useful to show, in addition to the mean signal, several realizations consistent with its probability distribution [Eq. (3.2) or (3.6)]. Looking at several realizations allows one to see which features are significant and which are not. Realizations can also be useful in the 1D case to make up for the fact that the confidence region does not contain any information about correlated uncertainties. For the applications here, though, we have not found them to be useful and so have not shown any.

IV. BAYESIAN COMPARISON

A natural question to ask is, ''How consistent are the two datasets?'' The Wiener filter gives a visual, qualitative answer to the question, but we would also like some quantitative answers as well. A better-formulated question is, ''Is my model of the data an adequate description of the two datasets together?'' To answer this question, one can extend the model of the data to include a residual and then check to see if this extension increases the likelihood. For example, one could add a residual that is Gaussian-distributed with zero mean:

$$\Delta_i = s_i + n_i + r_i$$

$$\langle \Delta_i \Delta_j \rangle = C_{T,ij} + C_{N,ij} + C_{\text{res},ij} .$$

$$(4.1)$$

Further restrictions on the form of $C_{\text{res},ij}$ must be made for the problem to not be degenerate. One could choose $C_{\text{res},ij}$ to be appropriate for a particular foreground contaminant [19], increased noise [12], or anything else that inspection of the data, combined with prior knowledge, has led the analyzer to suspect.

Below we describe a particular choice of $C_{\text{res},ij}$ that is useful in the absence of any hints as to the likely nature of a possible contaminant. One of its virtues is that it tests the agreement *between* the two datasets, rather than anything internal to them. This is because the residual can only be constrained by the comparison of the two datasets; the datasets individually cannot distinguish this residual from the signal itself. Another virtue of the residual model is that a statistic based on it has advantages over various χ^2 statistics, as is discussed in the next section. We point out though that discovering a residual as modeled below may be only the first step in uncovering the true nature of the residual. Once

a specific hypothesis is advanced, it could be tested with a $C_{\text{res},ij}$ more appropriate than the one described here.

A. The contamination parameter, γ

To test the consistency of the pairs of datasets—or rather, to test the adequacy of our model of the datasets—we introduce the following residual:

$$\Delta_1 = s_1 + n_1 + \gamma_1 r_1 \qquad (4.2)$$

and likewise for Δ_2. To reduce the number of parameters in this model for the residual, we set $\gamma = \gamma_1 = \gamma_2$. Now we must specify the probability distribution of r. For simplicity, let us take it to be a Gaussian random variable with zero mean. Clearly, we want the cross-term in the variance to be zero ($\langle r_1 r_2^\dagger \rangle = 0$), since we have in mind contaminants that are particular to each dataset. There is a lot of freedom in the choice of $\langle r_1 r_1^\dagger \rangle$ and $\langle r_2 r_2^\dagger \rangle$—once again for simplicity let us take these to be equal to C_{T11} and C_{T22}.

We have just added one parameter to whatever other parameters we were using to define the power spectrum. The model for the power spectrum we use here is standard cold dark matter (CDM), with the amplitude as the only free parameter. We have expressed the amplitude as σ_8—the *rms* fluctuations in mass in $8h^{-1}$ Mpc spheres. The experiments in question do not have sufficient dynamic range to constrain *strongly* more than this one parameter. For COBE-normalized standard CDM, $\sigma_8 = 1.2$.

We can now explicitly show the complete parameter dependence of the covariance matrix in our model, by modifying Eq. (3.5) to

$$C = \begin{pmatrix} \sigma_8^2 (1 + \gamma^2) \tilde{C}_{T11} + C_{N11} & \sigma_8^2 \tilde{C}_{T12} \\ \sigma_8^2 \tilde{C}_{T21} & \sigma_8^2 (1 + \gamma^2) \tilde{C}_{T22} + C_{N22} \end{pmatrix}$$

$$(4.3)$$

where the tilde means the quantity is evaluated for $\sigma_8 = 1$.

We prefer to work with a slightly different parameterization (spanning the same model space) by replacing σ_8^2 with $(\sigma_8')^2 \equiv \sigma_8^2 (1 + \gamma^2)$ which is the amplitude for the variance of the signal and the contaminant combined. We prefer σ_8' to σ_8 since its probability distribution of this quantity should be relatively independent of the level of contamination. Further, we prefer to use the fraction of contamination, $\gamma/\sqrt{(1 + \gamma^2)}$ rather than the contamination parameter itself. Probability distributions for σ_8' and $\gamma/\sqrt{(1 + \gamma^2)}$ can be seen in Fig. 8.

One can see from the shape of the contour curves that $\gamma/\sqrt{(1 + \gamma^2)}$ and σ_8' are very nearly uncorrelated. The reason is that the dominant contribution to the determination of σ_8 comes from terms in the likelihood proportional to $\Delta_i \Delta_j$ where Δ_i and Δ_j are in the same dataset, whereas γ is entirely determined by the cross-terms.

The most probable level of contamination indicated by the MSAM92/MSAM94 comparison is about 12%. However, there is virtually no evidence for non-zero contamination since the probability of zero contamination is only about 5% less. Complete contamination is strongly ruled out at more than $\exp(5^2/2) \simeq 2.7 \times 10^5$ times less probable. The

$\gamma/(1+\gamma^2)^{1/2}$ fractional contamination
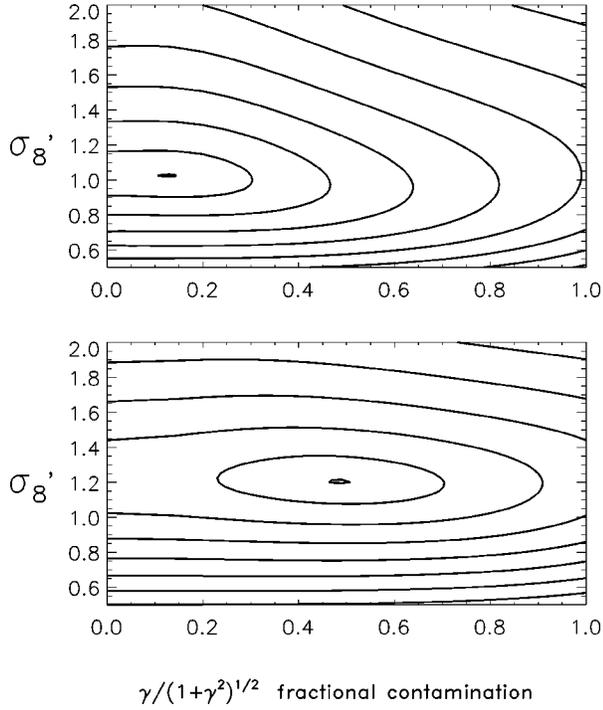
FIG. 8. Contours of the likelihood of σ_8' vs the fractional contamination for the MSAM92 and MSAM94 datasets (top panel) and the MSAM92 and SK95 datasets (bottom panel). The contours indicate reductions in probability from the maximum by factors of $e^{1^2/2}$, $e^{2^2/2}$, $e^{3^2/2}$, etc.

MSAM92/SK95 datasets are much less constraining on the amount of contamination that may be present. While 50% is the most probable value, total contamination and no contamination are only about 13 and 1.6 times less likely, respectively.

For simplicity, the contaminant is modeled as being uncorrelated between the two datasets and yet boosting the variance of each by the same amount. It is highly unlikely that a real contaminant would have this property. Again, if one suspects one experiment of being contaminated more than another, it is straightforward to change $C_{\text{res},ij}$ accordingly. We point out though that even as modeled here, a contaminant in just one of the experiments will still push the most likely γ away from zero. This will happen to accommodate the fact that the contaminant is not contributing to the dataset 1—dataset 2 cross-terms.

B. The probability enhancement factor, β

For many purposes, a much smaller extension into alternative hypothesis space may be useful. In particular, instead of examining a continuum, one could just compare the model with $\gamma=0$ to the model with $\gamma=\infty$, at fixed σ_8'. The interesting quantity is how much more probable one model is than the other, a quantity referred to as the odds. This particular odds, or rather its logarithm, we refer to as β and call it the probability enhancement factor:

$$\beta \equiv \ln \frac{P(\Delta_1\Delta_2|H_0)}{P(\Delta_1\Delta_2|H_\infty)} \qquad (4.4)$$

where H_0 (not to be confused with the present value of the Hubble constant) is the hypothesis that $\gamma=0$ and H_∞ is the hypothesis that $\gamma=\infty$. Both hypotheses are understood to be fixed at the same σ_8'. One can see from Eq. (4.3) that the cross-terms connecting the two different datasets in the covariance matrix C vanish when $\gamma\rightarrow\infty$ with σ_8' fixed. Therefore we can also write β as

$$\beta = \ln \frac{P(\Delta_1\Delta_2|C)}{P(\Delta_1|C)P(\Delta_2|C)} = \ln \frac{P(\Delta_1|\Delta_2,C)}{P(\Delta_1|C)} \qquad (4.5)$$

where C is understood to be C in Eq. (4.3) with $\gamma=0$ and the second equality follows from the use of $P(AB|C)=P(A|BC)P(B|C)$. This second equality gives rise to another interpretation of β: β indicates how much more probable dataset 1 is given that dataset 2 exists than it would be without the existence of dataset 2. And by the symmetry of the definition of β, we know that the statement is true under switching of 1 and 2.[3]

The probability enhancement factor, like the Wiener filter, depends on the assumed power spectrum used to calculate the theoretical covariance matrices. We find that for our parametrized model, within the most likely region of parameter space, the dependence of β on the parameter is weak. In Fig. 9 we see the dependence of $\beta(92,95)$ and $\beta(94,95)$ on σ_8.[4] Notice that this dependence is quite weak near the maximum of the joint and individual likelihoods. That β is minimized near the maximum of the likelihoods, is due to the fact that the joint likelihood [proportional to $P(\Delta_1\Delta_2|C)$] is broader than the product of the two individual likelihoods [proportional to $P(\Delta_1|C)P(\Delta_2|C)$]; see Eq. (4.5).

As can be seen from the log likelihood curves, the different datasets prefer slightly different values of σ_8.[5] For all calculations of β below and for the Wiener filtering in the previous section, we have chosen a value of $\sigma_8=1.2$, in between the preferred values for SK, MSAM92 and MSAM94. It is also the normalization for this power spectrum suggested by the DMR data.

V. FREQUENTIST STATISTICS

We now discuss β from the frequentist perspective. The frequentist approach to checking the consistency of a dataset is to invent some function of the data, called a statistic, and then to compare the measured value of the statistic to its probability distribution under various hypotheses. The prob-

---

[3]There is even a third interpretation of β as the log of the ratio of probability of no relative pointing error, to that of a gross relative pointing error which leaves the fields completely uncorrelated.

[4]To be precise, we mean σ_8', but in the following, we drop this prime for simplicity and also because keeping the prime does not make sense in the context of the interpretation of β as the increase in probability of one dataset given the other dataset.

[5]Some of this discrepancy may be due to calibration uncertainty which is not included in these log likelihood curves. We address this issue in a later section.
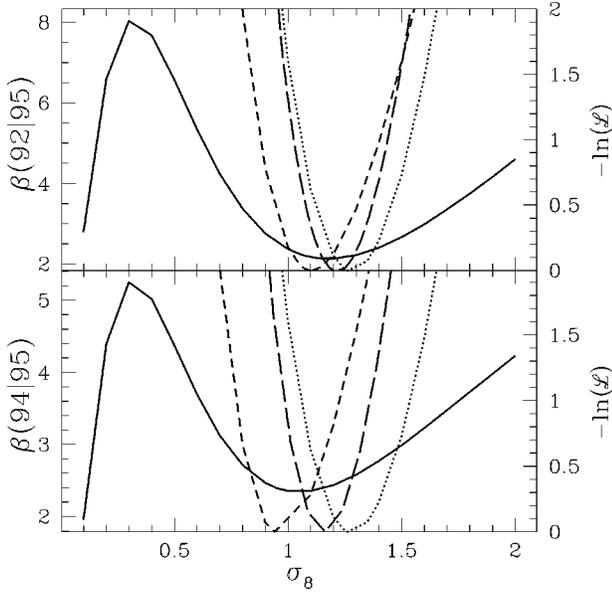
FIG. 9. Probability enhancement factor $\beta(92,95)$ (top panel) and $\beta(94,95)$ (bottom panel) as a function of σ_8 (solid curves). Also shown are $-\delta \ln \mathcal{L}$ for individual and joint datasets. Identifying these curves by their minima, they are, from left to right: MSAM92, MSAM92+SK, SK in the top panel and MSAM94, MSAM94+SK, SK in the bottom panel.

ability enhancement factor, β, can be viewed as a statistic since it is a function of the data. In fact, it is the logarithm of the well-known likelihood ratio statistic—in this case the ratio of the likelihood of H_0 to the likelihood of H_∞.

Some statistics are better than others at distinguishing among competing hypotheses. In this section, we see how β and other statistics fare at discriminating between hypotheses H_0 and H_∞.

A. Probability distributions of quadratic statistics

We restrict ourselves to studying quadratic functions of the data, for which we have analytic expressions for the mean and variance. In addition to various different χ^2 quantities (to be defined below), the probability enhancement factor—due to the logarithm in the definition—is also a quadratic function of the data:

$$\beta = (N/2)\ln|C| + \tfrac{1}{2}\Delta^T C^{-1}\Delta - (N_1/2)\ln|C_{11}| - \tfrac{1}{2}\Delta_1^T C_{11}^{-1}\Delta_1$$
$$- (N_2/2)\ln|C_{22}| - \tfrac{1}{2}\Delta_2^T C_{22}^{-1}\Delta_2, \qquad (5.1)$$

which follows from Eq. (4.5). Since it is a quadratic function of the data, it is straightforward to calculate the mean and variance.

In general, any quadratic function of the data, $Q \equiv \Delta^\dagger M \Delta + \text{constant}$, has a mean under hypothesis X of

$$\bar{Q}_X \equiv \langle Q \rangle_X = \text{Tr}(C_X M) + \text{const} \qquad (5.2)$$

and a variance of

$$\delta Q_X^2 \equiv \langle (\bar{Q}_X - Q)^2 \rangle_X = 2\,\text{Tr}(C_X M C_X M) \qquad (5.3)$$
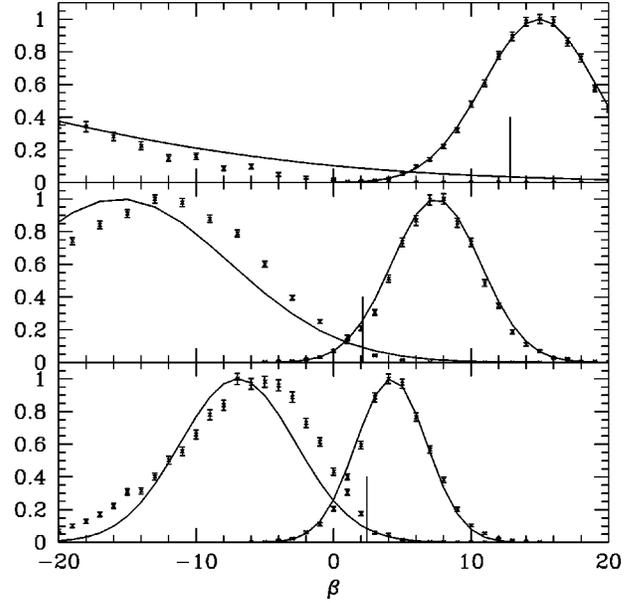
FIG. 10. The measured values of β (vertical lines) and its (arbitrarily normalized) probability distribution functions under the two hypotheses. From top to bottom: $\beta(92,94)$, $\beta(92,95)$ and $\beta(94,95)$. The curves peaking at positive β are estimates of $P(\beta|H_0)$ and those peaking at negative β are estimates of $P(\beta|H_\infty)$. The points with error bars are the results of a Monte Carlo calculation, while the solid curves are Gaussians with the analytically calculated means and variances.

where hypothesis X is specified by $C_X \equiv \langle \Delta\Delta^\dagger \rangle_X$.

For the case of β we have, for hypotheses H_0 and H_∞:

$$\langle \beta \rangle_0 = \frac{1}{2}\ln\left(\frac{|C_{11}|^{N_1}|C_{22}|^{N_2}}{|C|^N} \right)$$

$$\langle (\beta - \langle \beta \rangle_0)^2 \rangle_0 = \text{Tr}(w_{12}w_{21})$$

$$\langle \beta \rangle_\infty = \langle \beta \rangle_0 + \frac{1}{2}\,\text{Tr}(1 - C_\infty C^{-1})$$

$$\langle (\beta - \langle \beta \rangle_\infty)^2 \rangle_\infty = \frac{1}{2}\,\text{Tr}[(1 - C_\infty C^{-1})(1 - C_\infty C^{-1})]. \qquad (5.4)$$

Note that if the experiments have nothing to do with each other ($C_{12}=0$), then the numerator and denominator of the argument of the logarithm are equal and therefore $\langle \beta \rangle_0 = 0$ as we expect from the definition of β in Eq. (4.4).

Given the observed value of β, we can assess the validity of the two hypotheses by calculating the probability distribution of β under each hypothesis. As shown above, we can calculate the mean and variance analytically. To calculate the entire (non-Gaussian) distribution function though, we have used the Monte Carlo method. The results are plotted in Fig. 10 for the three possible pairings of the three datasets. The Monte Carlo method is quick because we first rotate to a basis where everything is diagonal and then make the realizations. The rotation to the diagonal basis only needs to be found once. The plots shown use between 4000 and 17000

realizations. Notice that the distribution of β under H_0 is well-approximated by a Gaussian. The deviations from Gaussianity are larger under H_∞.

We see in the figure that $\beta(92,94) = 13$ which is consistent with the expected range for hypothesis 0 of 15 ± 4.1. As a measure of the consistency, we have calculated the probability of getting a β greater than this to be 0.70. We also see that under hypothesis H_∞ such a value of β is extremely unlikely; the probability of getting a β greater than the measured one is less than 1%. We also find consistency with H_0 for the other two pairs of datasets: $\beta(92,95) = 2.1$ (c.f. $\langle\beta\rangle_0 = 7.4 \pm 3.2$) and $\beta(94,95) = 2.4$ (c.f. $\langle\beta\rangle_0 = 4.4 \pm 2.6$). For both of these, under hypothesis H^∞, the probability of getting a value of β as high or higher than the measured one is 1%.

B. Comparison of comparisons

There are a handful of other quadratic functions of the data one might consider using for comparison of datasets. Here we define the ones under consideration by specifying the data vectors on which they are based:

$$\chi_J^2 : \Delta \tag{5.5}$$

$$\chi_w^2 : \Delta - w\Delta \tag{5.6}$$

$$\chi_{w1}^2 : (\Delta_1 - w_{12}\Delta_2) \tag{5.7}$$

$$\chi_{w2}^2 : (\Delta_2 - w_{21}\Delta_1) \tag{5.8}$$

$$\chi_{w12}^2 : (w_{12}\Delta_2 - w_{11}\Delta_1) \tag{5.9}$$

$$\chi_{w21}^2 : (w_{21}\Delta_1 - w_{22}\Delta_2). \tag{5.10}$$

We clarify what we mean by two examples:

$$\chi_J^2 = \Delta^\dagger M^{-1} \Delta \tag{5.11}$$

where $M = \langle \Delta\Delta^\dagger \rangle_0 = C$, and

$$\chi_{w12}^2 = (w_{12}\Delta_2 - w_{11}\Delta_1)^\dagger M^{-1} (w_{12}\Delta_2 - w_{11}\Delta_1), \tag{5.12}$$

where

$$M \equiv \langle (w_{12}\Delta_2 - w_{11}\Delta_1)(w_{12}\Delta_2 - w_{11}\Delta_1)^\dagger \rangle_0$$
$$= (w_{11} - w_{12}w_{21})C_{T11} + (w_{12} - w_{11}w_{12})C_{T21}. \tag{5.13}$$

The J stands for joint, since this is the χ^2 quantity in the joint likelihood function, $P(\Delta|C)$. It is straightforward to show that $\chi_J^2 = \chi_w^2$, but, other than this relationship, the above χ^2s are all independent quantities.

To judge the discriminating power of all our quadratic statistics, we use the *separation factor*,

$$|Q_0 - Q_\infty|/\delta Q_0, \tag{5.14}$$
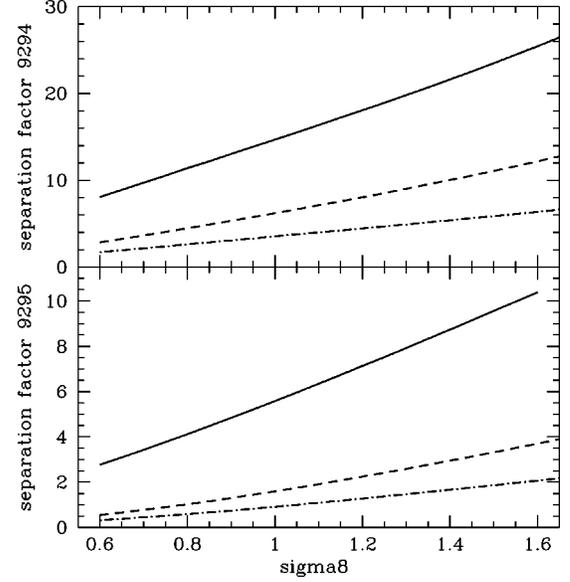
FIG. 11. Separation factors for β (blue, solid), χ_{w12}^2 (green, dashed) and χ_J^2 (magenta, dot-dashed). The top panel is for the 92/94 comparison and the bottom panel for the 92/95 comparison. For χ_{w12}^2 the smaller dataset is taken to be dataset 1.

where Q_0, Q_∞ are the means under the two hypotheses and δQ_0 is the standard error under H_0. The separation factor is shown as a function of σ_8 in Fig. 11. To avoid clutter, only two of the χ^2 quantities are shown, χ_J^2 and χ_{w12}^2. The separation factors for the other χ^2s are bounded by these two.

One can clearly see the superiority, under this measure, of the Bayesian-motivated probability enhancement factor. For example, for $\sigma_8 = 0.6$, if we assume H_0, it requires an 8σ fluctuation to get $\beta = \langle\beta\rangle_\infty$, but only a 2σ (3σ) fluctuation to get $\chi_J^2 = \langle\chi_J^2\rangle_\infty$ ($\chi_{w12}^2 = \langle\chi_{w12}^2\rangle_\infty$). The increase in all the separation factors with increasing σ_8 is expected since discriminating power should increase with increasing signal-to-noise of the measurements.

The separation factor, as we have defined it, is the separation between the expected value of the two hypotheses in units of the standard error assuming H_0 (δQ_0). One might also choose as another measure of discriminating power, this separation in units of the standard error assuming H_∞ (δQ_∞). In showing that β performs well under this measure, we are assisted by a theorem: the likelihood ratio test is *most powerful*.

A simple hypothesis test can be made from any statistic by choosing some critical value Q^*: if $Q > Q^*$, then reject H_0; otherwise, accept H_0.[6] Statisticians discuss the *size* and *power* of a test designed to discriminate between two hypotheses. The size of the test is the probability of rejecting H_0 if H_0 is true, while the power is the probability of rejecting H_0, if H_∞ is true. Clearly, we want the test to be such that the size is small and the power is large. By changing Q^*, we can choose the size. The test based on the likelihood

_____

[6]This assumes $Q_0 < Q_\infty$, if not then the test should be changed so that H_0 is rejected when $Q > Q^*$.

TABLE I. The probability enhancement factor is symmetric under the interchange of the two datasets, but χ^2_{w12} [defined in Eq. (5.9)] is not so we must specify that the datasets column has the format dataset 1, dataset 2.

Datasets	β	$\langle\beta\rangle_0 \pm \delta\beta_0$	$\langle\beta\rangle_\infty \pm \delta\beta_\infty$	χ^2_{w12}/ν	ν
92,94_2	5.7	10.9 ± 3.6	-37.6 ± 20.0	1.08	218
92,94_3	14.5	11.2 ± 3.6	-39.0 ± 20.1	1.05	218
92,94	12.8	15.0 ± 4.1	-58.4 ± 27.4	1.02	218
92,95_3	-2.5	4.4 ± 2.5	-8.5 ± 5.4	1.12	218
92,95_4	4.6	3.2 ± 2.2	-5.3 ± 3.6	1.11	218
92,95_5	-1.2	1.6 ± 1.7	-2.031 ± 1.4	1.05	218
92,95_6	-0.29	0.56 ± 1.03	-0.61 ± 0.49	1.06	218
92,95	2.13	7.4 ± 3.2	-15.6 ± 8.1	1.15	218
94,95_3	2.6	2.71 ± 2.08	-4.27 ± 3.05	1.02	170
94,95_4	1.4	1.94 ± 1.82	-2.69 ± 2.01	1.05	170
94,95_5	-0.31	0.99 ± 1.35	-1.14 ± 0.85	1.06	170
94,95_6	-0.99	0.35 ± 0.82	-0.365 ± 0.29	0.96	170
94,95	2.437	4.4 ± 2.63	-7.29 ± 4.2	1.05	170
92_2,92_3	8.29	8.80 ± 3.185	-31.6 ± 18.775	1.16	109
94_2,94_3	6.81	11.1 ± 3.418	-52.4 ± 30.4	0.93	85
95_3,95_4	1.72	5.2 ± 2.99	-7.12 ± 3.12	1.25	95
95_3,95_5	0.65	1.3 ± 1.59	-1.40 ± 0.62	1.09	95
95_3,95_6	1.29	0.39 ± 0.87	-0.40 ± 0.216	1.08	95
95_4,95_5	-1.03	2.14 ± 2.00	-2.46 ± 1.19	1.48	95
95_4,95_6	2.26	0.39 ± 0.88	-0.40 ± 0.18	1.05	95
95_5,95_6	-0.10	0.232 ± 0.68	-0.24 ± 0.13	1.08	95

ratio statistic has the property that, for a given size, it is most powerful—that is, no other test with the same size has a greater power. For a discussion of the likelihood ratio statistic in the context of CMB observations see, e.g. [20].

To see the relevance with our separation factor, let's set $Q^*=Q_0$. Let's further assume that we are in the asymptotic limit of large datasets so that all probability distributions are Gaussian. With $Q^*=Q_0$, the size of the test will be 0.5 for all statistics. Since the size of this test is the same for all statistics, we know that the likelihood ratio test ($Q=\beta$) will have the largest power. For $Q^*=Q_0$ the power is given by

$$\text{power}=1/2+\frac{1}{\sqrt{2\pi}\delta Q_\infty}\int_{Q_0}^{Q_\infty}\exp(Q-Q_\infty)^2/(2\,\delta Q^2_\infty)$$

$$=1/2+\text{erf}((Q_0-Q_\infty)/\sqrt{2\,\delta Q^2_\infty})/2. \quad (5.15)$$

Since the error function monotonically increases with its argument, we see that the separation between Q_0 and Q_∞ in units of δQ_∞ will always be largest for the likelihood ratio statistic, β.

We end this section with a brief consideration of one more χ^2 quantity. One could ask if there is a set of map pixels, T, that is consistent with the noise distribution:

$$\chi^2_n\equiv(\Delta-s)^\dagger C_n^{-1}(\Delta-s);\;\; s\equiv HT. \quad (5.16)$$

Because of its model independence, one might also think that χ^2_n is a compelling choice for testing the consistency of two datasets. However, if the pixels for the two datasets are

slightly different, then it will almost certainly be the case that a set of map pixels can be found that gives a reduced χ^2_n near unity. The problem is that this sky map may contain sharp spikes, highly inconsistent with our prior assumptions.

VI. APPLYING β TO SUBSETS OF DATA

We have also calculated β for various pairings of subsets of the data; the results are in Table I. All but one pairing (to be discussed later) have values of β within 2σ of $\langle\beta\rangle_0$. Note that the last 8 rows of the table are the results for internal consistency checks.

Also included in the table are the values of χ^2_{w12}. Under the separation factor criterion, this was the best other quadratic statistic. It is also of particular relevance to Figs. 5, 6 and 7, since these show the data vectors on which χ^2_{w12} is based.

Most of the reduced χ^2_{w12} values are comfortably close to unity. The probability of exceeding χ^2 is less than 5% for only one of the entries—the 95_4,95_5 internal consistency check for which the probability is less than 1%.

We have also found another breakup of the data to be useful. To identify localized problems in the data, we have calculated β as a function of the amount of data included. For example, in Fig. 12, we have plotted $\beta(92,95^*)$ vs α^*, where the star in 95* indicates that only 95 data with RA $\alpha<\alpha^*$ have been included. One can see here features associated with the discrepancies seen in the Wiener filter figures. Figures 13 and 14 show the results of similar calculations. For Fig. 14, the order in which the data is included is re-
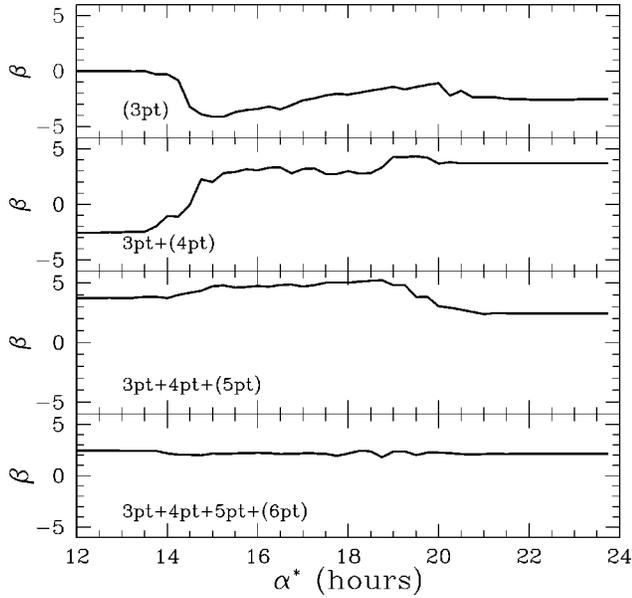
FIG. 12. The probability enhancement factor $\beta(92,95^*)$, where the* indicates that only data with right ascension, α, less than α^* are included. In the top panel, only the 3 point data are included for 95. In the panel one lower, in addition to all the 3 point data, those 4 point data with $\alpha < \alpha^*$ are included, etc.

FIG. 14. The probability enhancement factor, $\beta(92,94^*)$, where here* indicates that only data with right ascension, α, *greater* than α^* are included. In the top panel, only the 3-beam data are being included for 94. In the bottom panel, all the 3-beam data are included, but only those 2-beam data points with $\alpha > \alpha^*$. The dashed lines are $\langle \beta(92,94^*) \rangle_0$ and one standard deviation above and below.

versed (see caption) in order not to overemphasize the discrepant data at low RA.

From the first two entries of Table I and also from Fig. 14 we see that the 3-beam datasets are more consistent with each other than the 2-beam datasets, where there is a hint of a problem at low RA. This hint can also be seen in the

FIG. 13. The probability enhancement factor $\beta(95,92^*)$. The* indicates that only data with right ascension, α, less than α^* are included. In the top panel, only the 2-beam data are included for 92. In the bottom panel, in addition to all the 2-beam data, the 3-beam data with $\alpha < \alpha^*$ are included.

Wiener-filtered data shown in Fig. 5. Possibly confusing is that in Fig. 5 the discrepancy looks stronger in the 3-beam Wiener-filtered data. However, this is because the 2-beam data has significant influence on the best estimate of the 3-beam signal. Evidence for this relevance of the 2-beam data for the 3-beam data (and vice-versa) comes from the fact that $\beta(92\_2,92\_3)$ and $\beta(94\_2,94\_3)$ are large at 8.3 and 6.8, respectively. A further clue that the problem is with the 2-beam data is in Fig. 13 where there is a suggestion of a problem at low RA with the 2-beam, but not the 3-beam.

The better agreement between the 3-beam datasets than between the 2-beam datasets is possibly due to the greater susceptibility of the 2-beam data to atmospheric contamination. In particular, the 2-beam data is susceptible to atmospheric gradients, while the 3-beam is not. A gradient can arise as the pendulating motion of the gondola causes the motion of the chopping flat to be slightly different from constant elevation [21]. Presumably, one could test this hypothesis by searching for signals in the time stream with the balloon pendulation period.

Both from the Wiener filter figures and the cumulative β plot (Fig. 14) we can see that the MSAM92 and MSAM94 data agree very well at large RA and therefore what is observed is really on the sky and not some instrumental artifact. In contrast, the MSAM92/SK Wiener filter figures and cumulative β plots show discrepancies. These may be due to instrumental problems—in which case the problem must be with SK95—or foreground contamination which could affect either instrument. We discuss the possibility of foreground contamination in Sec. VIII on dust.

VII. FIXING CALIBRATIONS BY COMPARING DATASETS

We now give an example of searching for a particular "contaminant," relative calibration uncertainty. Every dataset must be calibrated by using the same apparatus to observe a radiation source of known brightness. This observation allows for the conversion of the data from some arbitrary units to temperature or brightness units. Often the brightness of the source in the passband of interest is only known to 10% or so in which case the calibration is a significant source of uncertainty. If Δ' is the uncalibrated data, then we define the calibration factor f as $\Delta = f\Delta'$, where Δ is the calibrated data. Similarly, the uncalibrated noise covariance matrix gets adjusted by f^2: $C_n = f^2 C'_n$, since the noise is determined from the data itself.

One can do likelihood analysis on the uncalibrated data, but with the appropriate covariance matrix:

$$C' \equiv \langle \Delta'(\Delta')^\dagger \rangle = \langle (s/f)(s/f)^\dagger \rangle + C'_n = \left(\frac{\sigma_8}{f}\right)^2 \widetilde{C}_T + C'_n. \tag{7.1}$$

For a joint dataset, Δ_1 and Δ_2, we have (dropping the primes):

$$C = \begin{pmatrix} \left(\dfrac{\sigma_8}{f_1}\right)^2 \widetilde{C}_{T11} + C_{n11} & \left(\dfrac{\sigma_8}{f_1}\right)\left(\dfrac{\sigma_8}{f_2}\right) \widetilde{C}_{T12} \\[2ex] \left(\dfrac{\sigma_8}{f_1}\right)\left(\dfrac{\sigma_8}{f_2}\right) \widetilde{C}_{T21} & \left(\dfrac{\sigma_8}{f_2}\right)^2 \widetilde{C}_{T22} + C_{n22} \end{pmatrix}.$$

Note that this covariance matrix, and hence the likelihood, depends only on σ_8/f_1 and σ_8/f_2. The degeneracy among the three parameters is broken by the calibration measurements of each experiment, which are usually modeled by a Gaussian:

$$\ln \mathcal{L}_{\text{tot}}(\sigma_8, f_1, f_2) = \ln \mathcal{L}\left(\frac{\sigma_8}{f_1}, \frac{\sigma_8}{f_2}\right) - \frac{(f_1 - \bar{f}_1)^2}{2\sigma_1^2} - \frac{(f_2 - \bar{f}_2)^2}{2\sigma_2^2}. \tag{7.2}$$

If one is exploring this likelihood space by direct evaluation, note that one can first evaluate $\ln \mathcal{L}$ on a two dimensional grid (σ_8/f_1, σ_8/f_2) and then evaluate the three-dimensional $\ln \mathcal{L}_{\text{tot}}$ by adding in the (very easy to calculate) calibration measurement terms. The data as we receive it has already been calibrated so we usually take $\bar{f} = 1$.

We have evaluated this $\ln \mathcal{L}_{\text{tot}}$ for the MSAM92 dataset and a subset of the SK95 ring dataset, from 13^h to 22^h, covering the range of influence of the MSAM92 data. For MSAM92 we take $\bar{f}_{\text{MSAM92}} = 1$ and $\sigma_{\text{MSAM92}} = 0.1$. For SK we use the Netterfield *et al.* calibration $\bar{f}_{\text{SK}} = 1$ and $\sigma_{\text{SK}} = 0.14$. We find in this case that $\ln \mathcal{L}_{\text{tot}}$ is minimized at $f_{\text{MSAM92}} = 0.99$, $f_{\text{SK}} = 0.99$ and $\sigma_8 = 1.13$. If the Leitch recalibration is used ($\bar{f}_{\text{SK}} = 1.05$, $\sigma_{\text{SK}} = 0.07$), then $\ln \mathcal{L}_{\text{tot}}$ is minimized at $f_{\text{MSAM92}} = 1.02$, $f_{\text{SK}} = 1.02$ and $\sigma_8 = 1.13$.
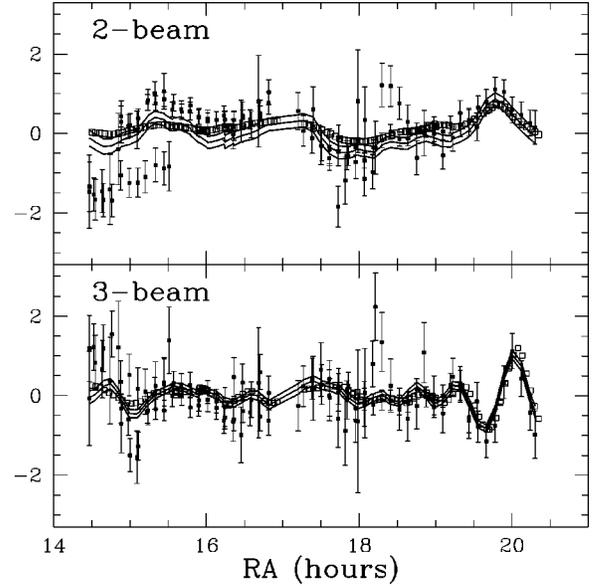
FIG. 15. Wiener-filtered dust. The points with error bars are the MSAM92 pixelized dust data. Two-beam in top panel, 3-beam in bottom panel. The three curves in each panel are the Wiener-filtered data bounded by $\pm$ one standard deviation (assuming Gaussianity). The open squares are the results of convolving the IRAS data with the MSAM beam pattern; the scale is set by a fit to the MSAM92 data.

We can also neglect the calibration measurements entirely and use the two datasets themselves to find the best relative calibration, $f_{12} \equiv f_1/f_2$.

$$\mathcal{L}_{12}(f_{12}) \propto \int dx_2 \mathcal{L}(x_1, x_2) \delta(x_1 - f_{12} x_2) \tag{7.3}$$

where $x_1 = f_1/\sigma_8$ and $x_2 = f_2/\sigma_8$. We find that, once again restricting the SK95 data to between 13 and 22 h that $f_{\text{MSAM92,SK}} = 1.06^{+.22}_{-.26}$. Netterfield *et al.* [5] find from their analysis that $f_{\text{MSAM92,SK}} = 1.22 \pm 0.24$.

Note that there is a possible problem for joint power spectrum analysis if relative calibration uncertainty is not taken into account. For overlapping experiments, neglect of this uncertainty could artificially boost high frequency power.

VIII. DUST

Via the application of β and Wiener-filtering, we have discovered a marginally significant discrepancy between the MSAM92 observations and those of Saskatoon at large RA. In this section we investigate the possibility that the discrepancy is due to foreground contamination of either the SK or MSAM92 datasets. The foreground contamination hypothesis is supported by the fact that the discrepancy occurs where the observations are closest to the plane of the galaxy. Further, from the MSAM92 interstellar dust data, one can see that the discrepancy occurs roughly where the dust is brightest—see Fig. 15.

Besides different spectral dependence from the CMB, interstellar dust also has a different spatial frequency dependence than the CMB. Gautier *et al.* [22], using the 100 mi-

cron IRAS data, and Wright [23], using the DIRBE data, both found that, away from the plane of the galaxy, the dust has an angular power spectrum, $C_\ell \propto \ell^{-3}$. Schlegel, Finkbeiner, and Davis [24], using a dust map produced by combining DIRBE and IRAS data, have inferred a slightly less steep spectrum with $C_\ell \propto \ell^{-2.5}$. For Fig. 15 we have used a power spectrum with $C_\ell \propto \ell^{-2.5}$ to Wiener filter the MSAM92 dust data. Using $C_\ell \propto \ell^{-3}$ produces very similar results. The dust is known to be highly non-Gaussian. While the mean signal does not depend on the statistics of the signal, the uncertainties in the signal do. Therefore one should bear this in mind when interpreting the graph, since the error bars in the figure were calculated assuming Gaussianity. Along with the MSAM92 dust data is the result of convolving the MSAM beam pattern with the IRAS SISSA 240 micron map [25]. The IRAS data have been scaled to fit the MSAM92 data. Note that the agreement for the 3-beam data is much better than for the 2-beam data, once again suggesting that it is a more reliable dataset.

What we have referred to as the MSAM92 CMB and dust data are obtained by fitting each pixel of the four frequency channels (170 GHz, 220 GHz, 500 GHz, 680 GHz) of MSAM92 data to a CMB component and a dust component. From this fit we get the CMB temperature and dust optical depth. The dust is assumed to be a ''gray'' body with temperature $T = 20$ K and emissivity index $\alpha = 1.5$.

Using this model, the dust feature at large RA should not be showing up in the lowest frequency channel. Therefore the fit ascribes the structure seen in the 170 GHz channel to CMB. However, the model may be an inadequate description; there may be a component correlated with the dust with stronger emission at 170 GHz than the thermal dust emission itself. Indeed, the shape of the dust feature at large RA is somewhat similar to the MSAM92 CMB feature at large RA.

The Saskatoon data is single frequency and thus harder to check directly for contamination. Despite the low frequency, dust (or a source correlated with dust) contamination is a possibility. Several datasets point to a correlation between high frequency maps dominated by thermal emission from dust and lower frequency measurements [26]. A weak, but significant, correlation has been seen [27,19] in a correlation analysis of the entire SK dataset with DIRBE and IRAS dust maps. Although weak, the correlation is measured by averaging over all the SK data; the effect may be large in certain regions, such as the one in question here near 18 h. The cause of this correlation is not yet known, although an hypothesis has been advanced by Draine and Lazarian [28] that it is due to dipole emission from spinning dust grains.

To investigate this possibility, we have Wiener-filtered the MSAM92 dust measurements onto the SK95 data in the region of overlap with MSAM92 (see Fig. 16). For the shape of the dust power spectrum, we used $C_\ell \propto \ell^{-2.5}$ [24] with amplitude chosen to maximize the likelihood given the MSAM92 dust data. Although, as expected, the dust is brightest in the region of the discrepancy, we have been unable to identify any more detailed relation between the predicted dust signal and the discrepancy.

There is another reason, besides the observed dust correlation, for believing the problem may lie with the SK data.
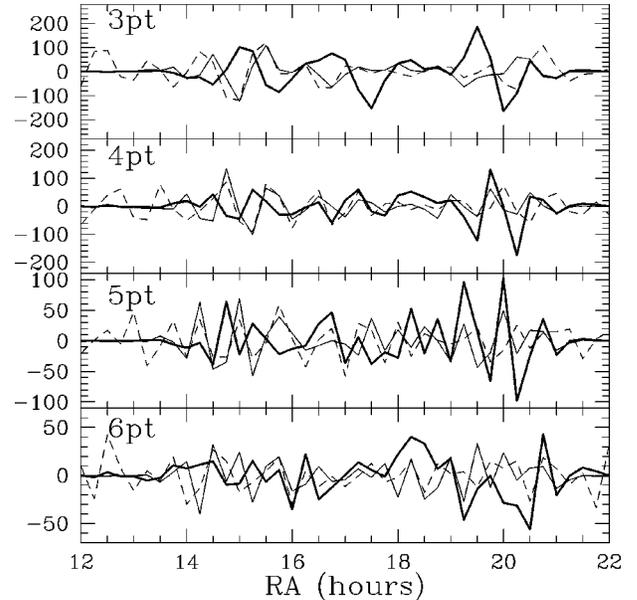
FIG. 16. The MSAM92 dust data (heavy solid line), MSAM92 CMB data (light solid line) and SK95 data all Wiener-filtered on to the SK95 pixels.

The SK team also did some internal consistency checks on their data, one of which is the A-B test. Their A and B detectors measured orthogonal linear polarizations, and thus for an unpolarized, or weakly polarized, signal, A-B should be consistent with zero. However, for the region of overlap with MSAM92, they find $\chi^2_{A-B}/\nu = 1.55$ for 80 degrees of freedom. The origin of this asymmetry is unclear, possibly an instrumental artifact. It could possibly be explained by rotational emission from dust grains; Draine and Lazarian predict that this component of the dust emission is between 0.1% and 10% polarized.

IX. SUMMARY

We have demonstrated the usefulness of the Wiener filter for making visual comparisons of datasets. We have emphasized that meaningful consistency testing requires alternative models with which to compare. Thus we have explicitly extended our model of the data to include a possible contaminant and calculated the probability distribution of the amplitude of this contaminant. For purposes of extracting just one number from the comparison, we advocate calculating the ratio of the probability of no contamination to the probability of infinite contamination. Viewed as a statistic, we have shown this ''probability enhancement factor'' to be better than various χ^2 statistics at discriminating between competing hypotheses.

The utility of our comparison statistics was shown by exercising them on the MSAM92, MSAM94 and SK95 data. We have found from comparing MSAM92 and MSAM94 that the most probable level of contamination is 12%, with zero contamination only 1.05 times less probable, and total contamination over 2×10^5 times less probable. From comparing MSAM92 and SK95, we have found that the most probable level of contamination is 50%, with zero contami-

nation only 1.6 times less probable, and total contamination 13 times less probable. Looking at subsets of the data, we find a region at large RA where the SK and MSAM92 measurements disagree. From IRAS and from the MSAM92 dust measurements, we know that this region is also the dustiest region of the overlap between SK and MSAM92. The origin of the discrepancy is unclear and may be due to instrumental artifacts in SK, or foreground contamination of either the SK or MSAM92 measurements.

A revolution is underway in the quality and quantity of CMB data—a revolution generated by the satellites MAP and Planck [29] as well as by a number of balloon and ground-based programs. The amount of data may soon be too large for the type of complete statistical analysis described here. However, any approximate methods developed for extracting the power spectrum or parameters will also be applicable to the statistical procedures introduced here.

[1] J. Mather *et al.*, Astrophys. J. Lett. **354**, L37 (1990); D. J. Fixsen *et al.*, *ibid.* **473**, 576 (1996).

[2] L. Knox, Phys. Rev. D **52**, 4307 (1995); G. Jungman, M. Kamionkowski, A. Kosowsky, and D. N. Spergel, Phys. Rev. Lett. **76**, 1007 (1996); Phys. Rev. D **54**, 1332 (1996); J. R. Bond, G. Efstathiou, and M. Tegmark, astro-ph/9702100; M. Zaldarriaga, D. Spergel, and U. Seljak, astro-ph/9702157.

[3] K. Ganga, E. Cheng, S. Meyer, and L. Page, Astrophys. J. Lett. **410**, L57 (1993).

[4] C. H. Lineweaver *et al.*, Astrophys. J. **448**, 482 (1995).

[5] C. B. Netterfield, M. J. Devlin, N. Jarosik, L. Page, and E. J. Wollack, Astrophys. J. **474**, 47 (1997); E. J. Wollack, M. J. Devlin, N. Jarosik, C. B. Netterfield, L. Page, and D. Wilkinson, *ibid.* **476**, 440 (1997).

[6] J. E. Ruhl, M. Dragovan, S. R. Platt, J. Kovac, and G. Novak, Astrophys. J. Lett. **453**, L1 (1995); S. R. Platt, J. Kovac, M. Dragovan, J. B. Peterson, and J. E. Ruhl, Astrophys. J. Lett. **475**, L1 (1994).

[7] C. A. Inman, E. S. Cheng, D. A. Cottingham, D. J. Fixsen, M. S. Kowitt, S. S. Meyer, L. A. Page, J. L. Puchalla, J. E. Ruhl, and R. F. Silverberg, Astrophys. J. Lett. **478**, L1 (1997).

[8] J. R. Bond, A. H. Jaffe, and L. Knox, Phys. Rev. D **57**, 2117 (1998).

[9] For a review of Bayesian inference, see Sec. 2 of P. C. Gregory and T. J. Loredo, Astrophys. J. **398**, 146 (1992); see also D. S. Sivia, *Data Analysis: A Bayesian Tutorial* (Clarendon, Oxford, 1996).

[10] For a more detailed derivation see, e.g., E. F. Bunn, Y. Hoffman, and J. Silk, Report No. astro-ph/9509045.

[11] M. Tegmark, A. de Oliveira-Costa, M. J. Devlin, C. B. Netterfield, L. Page, and E. J. Wollack, Astrophys. J. Lett. **474**, L77 (1997).

[12] J. R. Bond, Phys. Rev. Lett. **74**, 4369 (1994); J. R. Bond, in *Cosmology and Large Scale Structure, Les Houches Session LX*, edited by R. Schaeffer *et al.* (Elsevier, Amsterdam, 1996), pp. 469–674.

[13] D. J. Fixsen *et al.*, Astrophys. J. **470**, 63 (1996).

[14] E. S. Cheng *et al.*, Astrophys. J. Lett. **422**, L37 (1994).

[15] E. S. Cheng *et al.*, Astrophys. J. Lett. **456**, L71 (1996).

[16] E. S. Cheng *et al.*, Astrophys. J. Lett. **488**, 59 (1997); astro-ph/9705041.

[17] E. Leitch, Caltech Ph.D thesis, 1998, http://astro.caltech.edu/~eml/thesis.html

[18] C. B. Netterfield (private communication).

[19] A. H. Jaffe, D. Finkbeiner, and J. R. Bond (unpublished).

[20] Readhead *et al.*, Astrophys. J. **346**, 566 (1989).

[21] S. S. Meyer (private communication).

[22] T. N. I. Gautier, F. Boulanger, M. Perault, and J. L. Puget, Astron. J. **103**, 1313 (1992).

[23] E. L. Wright, Report No. astro-ph/9711261, 1997.

[24] D. J. Schlegel, D. P. Finkbeiner, and M. Davis, Astrophys. J. (to be published), astro-ph/9710327.

[25] J. Puchalla (private communication).

[26] A. Kogut *et al.*, Astrophys. J. Lett. **464**, L5 (1996); E. M. Leitch, A. C. S. Readhead, T. J. Pearson, and S. T. Myers, Report No. astro-ph/9705241.

[27] A. de Oliveira-Costa, A. Kogut, M. J. Devlin, C. B. Netterfield, L. A. Page, and E. J. Wollack, Astrophys. J. Lett. **482**, L17 (1997).

[28] B. Draine and A. Lazarian, Astrophys. J. Lett. **494**, 19 (1998).

[29] C. Bennett *et al.*, MAP experiment home page, http://map.gsfc.nasa.gov, 1996; M. Bersanelli *et al.*, *COBRAS/SAMBA, The Phase A Study for an ESA M3 Mission*, report, 1996; Planck home page, http://astro.estec.esa.nl/SA-general/Projects/Cobras/cobras.html, 1996.