March 2, 1994 CfPA 93-th-40

Power Spectrum Independent Constraints on Cosmological Models †

Max Tegmark
Emory Bunn
Wayne Hu
Department of Physics, University of California, Berkeley, California 94720

ABSTRACT

A formalism is presented that allows cosmological experiments to be tested for consistency, and allows a simple frequentist interpretation of the resulting significance levels. As an example of an application, this formalism is used to place constraints on bulk flows of galaxies using the results of the microwave background anisotropy experiments COBE and SP91, and a few simplifying approximations about the experimental window functions. It is found that if taken at face value, with the quoted errors, the recent detection by Lauer and Postman of a bulk flow of 689 km/s on scales of $150h^{-1}$Mpc is inconsistent with SP91 at a 95% confidence level within the framework of a Cold Dark Matter (CDM) model. The same consistency test is also used to place constraints that are completely model-independent, in the sense that they hold for any power spectrum whatsoever — the only assumption being that the random fields are Gaussian. It is shown that the resulting infinite-dimensional optimization problem reduces to a set of coupled non-linear equations that can readily be solved numerically. Applying this technique to the above-mentioned example, we find that the Lauer and Postman result is inconsistent with SP91 even if no assumptions whatsoever are made about the power spectrum.

† First submitted to Ap. J. in November 1993, revised in March 1994.

1. INTRODUCTION

Together with the classical cosmological parameters h, Ω, *etc.*, the power spectrum $P(k)$ of cosmological density fluctuations is one of the most sought-for quantities in modern cosmology, vital for understanding both the formation of large-scale structure and the fluctuations in the cosmic microwave background radiation (CMB).

The traditional approach has been to assume some functional form for $P(k)$ (like that predicted by the cold dark matter (CDM) scenario, for instance), and then investigate whether the predictions of the model are consistent with experimental data or not. The large amounts of data currently being produced by new CMB experiments and galaxy surveys, all probing different parts of the power spectrum, allow a new and more attractive approach. We can now begin to probe exact shape of the function $P(k)$, without making any prior assumptions about $P(k)$. More specifically, we measure different weighted averages of the function, the weights being the experimental window functions.

This new approach is quite timely (Juszkiewicz 1993), as there are now many indications that the primordial power spectrum may have been more complicated than an $n = 1$ power law. There are several sources of concern about the standard CDM cosmology, with inflation leading to $\Omega \approx 1$ and a primordial $n \approx 1$ Harrison-Zel'dovich power spectrum. Compared to COBE-normalized CDM, observational data shows unexpected large-scale bulk flows (Lauer & Postman 1993), too weak density correlations on small scales (Maddox *et al.* 1990), a rather quiet local velocity field (Schlegel *et al.* 1993) and a deficit of hot x-ray clusters (Oukbir & Blanchard 1992). The combined data from the COBE DMR (Smoot *et al.* 1992) and the Tenerife anisotropy experiment (Hancock *et al.* 1994) point to a spectral index exceeding unity (Watson & Gutiérrez de la Cruz 1993) which, if correct, cannot be explained by any of the standard inflationary models. The recent possible detections of halo gravitational microlensing events (Alcock *et al.* 1993) give increased credibility to the possibility that the dark matter in our galactic halo may be baryonic. If this is indeed the case, models with $\Omega < 1$ and nothing but baryonic dark matter (BDM) (Peebles 1987, Gnedin & Ostriker 1992, Cen, Ostriker & Peebles 1993) become rather appealing. However, in contrast to CDM with inflation, BDM models do not include a physical mechanism that makes a unique prediction for what the primeval power spectrum should be. Rather, the commonly assumed $P(k) \propto k^{-1/2}$ is chosen *ad hoc* to fit observational data. Moreover, for fluctuations near the curvature scale in open universes, where the $\Omega = 1$ Fourier modes are replaced by hyperspherical Bessel functions with the curvature radius as a built-in length scale, the whole notion of scale-invariance loses its meaning (Kamionowski & Spergel 1993).

In summary, it may be advisable to avoid theoretical prejudice as to the shape of the primordial power spectrum. In this spirit, we will develop a consistency test that requires no such assumptions whatsoever about the form of the power spectrum. This approach was pioneered by Juszkiewicz, Górski and Silk (1987), who developed a formalism for comparing two experiments in a power-spectrum independent manner. We generalize this method to the case of more than two experiments, and then use the formalism to assess the consistency of three recent observational results: the CMB anisotropy measurements made by the COBE Differential Microwave Radiometer (Smoot *et al.* 1992), the South Pole anisotropy experiment (SP91, Gaier *et al.* 1992), and the measurement of bulk velocity of Abell clusters in a 150 h^{-1} Mpc sphere (Lauer and Postman 1993, hereafter LP).

In Section 2, we develop a formalism for testing cosmological models for consistency. In Section 3, we apply this formalism to the special case of cold dark matter (CDM) and the LP, SP91 and COBE experiments. In Section 4, we solve the variational problem that arises in consistency tests of models where we allow arbitrary power-spectra, and apply these results to the LP, SP91 and COBE experiments. Section 5 contains a discussion of our results. Finally, two different goodness-of-fit parameters are compared in Appendix A, and the relevant window functions are derived in Appendix B.

2. CONSISTENCY TESTS FOR COSMOLOGICAL MODELS

In cosmology, a field where error bars tend to be large, conclusions can depend crucially on the probabilistic interpretation of confidence limits. Confusion has sometimes arisen from the fact that large-scale measurements of microwave background anisotropies and bulk flows are fraught with two quite distinct sources of statistical uncertainty, usually termed experimental noise and cosmic variance. In this section, we present a detailed prescription for testing any model for consistency with experiments, and discuss the appropriate probabilistic interpretation of this test. By *model* we will mean not merely a model for the underlying physics, which predicts the physical quantities that we wish to measure, but also a model for the various experiments. Such a model is allowed to contain any number of free parameters. In subsequent sections, we give examples of both a very narrow class of models (standard CDM where the only free parameter is the overall normalization of the power spectrum), and a wider class of models (gravitational instability with Gaussian adiabatic fluctuations in a flat universe with the standard recombination history, the power spectrum being an arbitrary function).

Suppose that we are interested in N physical quantities $c_1, ..., c_N$, and have N experiments $E_1, ..., E_N$ devised such that the experiment E_i measures the quantity c_i. Let s_i denote the number actually obtained by the experiment E_i. Because of experimental noise, cosmic variance, *etc.*, we do not expect s_i to exactly equal c_i. Rather, s_i is a random variable that will yield different values each time the experiment is repeated. By repeating the experiment M times on this planet and averaging the results, the uncertainty due to experimental noise can be reduced by a factor $\sqrt{M}$. However, if the same experiment were carried out in a number of different horizon volumes throughout the universe (or, if we have ergodicity, in an ensemble of universes with different realizations of the underlying random field), the results would also be expected to differ. This second source of uncertainty is known as cosmic variance. We will treat both of these uncertainties together by simply requiring the model to specify the probability distribution for the random variables s_i.

Let us assume that the random variables s_i are all independent, so that the joint probability distribution is simply the product of the individual probability distributions, which we will denote $f_i(s)$. This is an excellent approximation for the microwave background and bulk flow experiments we will consider. Finally, let $\hat{s}_1, ..., \hat{s}_N$ denote the numbers actually obtained in one realization of the experiments.

The general procedure for statistical testing will be as follows:

- First, define a parameter η that is some sort of measure of how well the observed data s_i agree with the probability distributions f_i, with higher η corresponding to a better fit.
- Then compute the probability distribution $f_\eta(\eta)$ of this parameter, either analytically or by employing Monte-Carlo techniques.
- Compute the observed value of η, which we will denote $\hat{\eta}$.
- Finally, compute the probability $P(\eta < \hat{\eta})$, *i.e.* the probability of getting as bad agreement as we do or worse.

We will now discuss these four steps in more detail.

2.1. Choosing a goodness-of-fit parameter

Obviously, the ability to reject models at a high level of significance depends crucially on making a good choice of goodness-of-fit parameter η. In the literature, a common choice is the *likelihood product*, *i.e.*

$$\eta_l \propto \prod_{i=1}^{N} f_i(s_i). \tag{1}$$

In this paper, we will instead use the *probability product*, *i.e.* the product of the probabilities P_i that each of the experiments yield results at least as extreme as observed. Thus if the observed $\hat{s}_i$ is smaller than the median of the distribution f_i, we have $P_i = 2P(s_i < \hat{s}_i)$, whereas $\hat{s}_i$ larger than the median would give $P_i = 2P(s_i > \hat{s}_i)$. The factor of two is present because we want a two-sided test. Thus $P_i = 1$ if $\hat{s}_i$ equals the median, $P_i = 2\%$ if $\hat{s}_i$ is at the high 99th percentile, *etc.*

These two goodness-of-fit parameters are compared in Appendix A for a few explicit examples, and the conclusions can be summarized as follows:

1) Rather unphysical probability distributions can be concocted that "fool" either one of these two parameters but not the other. Thus neither the likelihood product nor the probability product can be hailed as fundamentally better than the other.
2) For probability distributions encountered in the type of cosmology applications discussed in this paper, always smooth and unimodal functions, the likelihood product and the probability product yield very similar results.
3) As described in the following sections, the probability distribution of the probability product can always be computed analytically. The probability distribution for the likelihood product, however, depends on the distributions of the underlying random variables, and except in a few fortuitous simple cases, it must be computed numerically through either repeated convolutions or Monte Carlo simulations. Thus the probability product is considerably more convenient to use.

Since 1) and 2) indicate that neither of the two goodness-of-fit parameters is preferable over the other on scientific grounds, the authors feel that 3) tips the balance in favor of the probability product.

2.2. Its probability distribution

Apart from the simple interpretation of the probability product η, it has the advantage that its probability distribution can be calculated analytically, and is completely independent of the physics of the model — in fact, it depends only on N. We will now give the exact distributions.

By construction, $0 \leq \eta \leq 1$. For $N = 1$, η will simply have a uniform distribution, *i.e.*

$$f_\eta(\eta) = \begin{cases} 1 & \text{if } 0 \leq \eta \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Thus in the general case, η will be a product of N independent uniformly distributed random variables. The calculation of the probability distribution for η is straightforward, and can be found in a number of standard texts. The result is

$$f_\eta(\eta) = -f_z(-\ln\eta)\frac{dz}{d\eta} = \begin{cases} \frac{1}{(N-1)!}(-\ln\eta)^{N-1} & \text{if } 0 \leq \eta \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

2.3. The consistency probability

The probability $P(\eta < \hat{\eta})$, the probability of getting as bad agreement as we do or worse, is simply the cumulative distribution function $F_\eta(\hat{\eta})$, and the integral can be carried out analytically for any N:

$$F_\eta(\hat{\eta}) \equiv P(\eta < \hat{\eta}) = \int_0^{\hat{\eta}} f_\eta(u)du = \hat{\eta}\theta(\hat{\eta})\sum_{n=0}^{N-1}\frac{(-\ln\hat{\eta})^n}{n!}, \tag{4}$$

where θ is the Heaviside step function, and $F_\eta(\hat{\eta}) = 1$ for $\hat{\eta} \geq 1$. Since the product of N numbers between zero and one tends to zero as $N \to \infty$, it is no surprise that

$$F_\eta(\hat{\eta}) \to \theta(\hat{\eta})\hat{\eta}e^{-\ln\hat{\eta}} = \theta(\hat{\eta}) \tag{5}$$

as $N \to \infty$, *i.e.* that $f_\eta(\hat{\eta}) \to \delta(\hat{\eta})$. The function $F_\eta(\hat{\eta})$ is plotted in Figure 1, and the values of $\hat{\eta}$ for which $F_\eta(\hat{\eta}) = 0.05$, 0.01 and 0.001, respectively, are given in Table I for a few N-values. For example, if three experimental results give a goodness-of-fit parameter $\hat{\eta} = 0.0002$ for some model, then this model is ruled out at a confidence level of 99%. Thus if the model where true and the experiments where repeated in very many different horizon volumes of the universe, such a low goodness-of-fit value would be obtained less than 1% of the time.

Table 1. Probability product limits:

Confidence level	$N=1$	$N=2$	$N=3$	$N=4$
95%	0.05	0.0087	0.0018	0.00043
99%	0.01	0.0013	0.00022	0.000043
99.9%	0.001	0.000098	0.000013	0.0000021

2.4. Ruling out whole classes of models

If we wish to use the above formalism to test a whole set of models, then we need to solve an optimization problem to find the one model in the set for which the consistency probability is maximized. For instance, if the family of models under consideration is standard $n=1$, $\Gamma=0.5$ CDM (see Section 3), then the only free parameter is the overall normalization constant A. Thus we can write the consistency probability as $p(A)$, and use some numerical method to find the normalization A_* for which $p(A)$ is maximized. After this, the statistical interpretation is clear: if the experiments under consideration are carried out in an ensemble of CDM universes, as extreme results as those observed will only be obtained at most a fraction $p(A_*)$ of the time, whatever the true normalization constant is. Precisely this case will be treated in the next section. For the slightly wider class of models consisting of CDM power spectra with arbitrary A, n and Γ, the resulting optimization problem would be a three-dimensional one, and the maximal consistency probability would necessarily satisfy

$$p(A_*, n_*, \Gamma_*) \geq p(A_*, 1, 0.5) = p(A_*). \tag{6}$$

An even more general class of models is the set of all models where the random fields are Gaussian, *i.e.* allowing completely arbitrary power spectra P. In section 4, we will show that the resulting infinite-dimensional optimization problem can in be reduced to a succession of two finite-dimensional ones.

3. COLD DARK MATTER CONFRONTS SP91, COBE AND LAUER-POSTMAN

As an example of an application of the formalism presented in the previous section, we will now test the standard cold dark matter (CDM) model of structure formation for consistency with the SP91 CMB experiment and the Lauer-Postman bulk flow experiment.

Let E_1 be the Lauer-Postman (LP for short) measurement of bulk flows of galaxies in a $150h^{-1}$Mpc sphere (Lauer and Postman, 1993). Let E_2 be the 1991 South Pole CMB anisotropy experiment, SP91 for short (Gaier *et al.* 1992). Let E_3 be the COBE DMR experiment (Smoot *et al.* 1992). All of these experiments probe scales that are well described by linear perturbation theory, and so as long as the initial fluctuation are Gaussian, the expected results of the experiments can be expressed simply as integrals over the power spectrum of the matter perturbation:

$$\langle s_i \rangle = \int W_i(k) P(k) dk.$$

Here s_{sp} and s_{cobe} are the mean-square temperature fluctuations measured by the experiments, and $s_{lp} \equiv (v/c)^2$ is the squared bulk flow. The corresponding window functions W_{lp}, W_{sp} and W_{cobe} are derived in Appendix B, and plotted in Figure 2. These window functions assume that the initial perturbations were adiabatic, that $\Omega = 1$, and that recombination happened in the standard way, i.e. a last-scattering surface at $z \approx 1000$. The SP91 window function is to be interpreted as a lower limit to the true window function, as it includes contributions only from the Sachs-Wolfe effect, not from Doppler motions or intrinsic density fluctuations of the surface of last scattering.

Now let us turn to the probability distributions for the random variables s_{lp}, s_{sp} and s_c. The standard CDM model with power-law initial fluctuations $\propto k^n$ predicts a power spectrum that is well fitted by (Bond and Efstathiou 1984)

$$P(k) = \frac{Aq^n}{\left(1 + [aq + (bq)^{1.5} + (cq)^2]^{1.13}\right)^{2/1.13}}, \tag{7}$$

where $a = 6.4$, $b = 3.0$, $c = 1.7$ and $q = (1h^{-1}\mathrm{Mpc})k/\Gamma$. For the simplest model, $\Gamma = h$, but certain additional complications such as a non-zero cosmological constant Λ and a non-zero fraction Ω_ν of hot dark matter can be fitted with reasonable accuracy by other values of Γ (Efstathiou, Bond and White 1992). Thus the model has three free parameters: n, Γ and the overall normalization A. Integrating the power spectrum against the three window functions yields the values of c_i given in Table 2. The two rightmost columns contain the quotients c_{lp}/c_{sp} and c_{lp}/c_c, respectively.

Table 2. Expected r.m.s. signals for CDM power spectrum with $A = (1h^{-1}\mathrm{Mpc})^3$:

n	Γ	LP	SP91	COBE	LP/SP91	LP/COBE
1	0.5	9.2×10^{-7}	1.6×10^{-8}	2.0×10^{-8}	56.7	45.1
0.7	0.5	1.7×10^{-6}	2.9×10^{-8}	5.1×10^{-8}	57.2	32.7
2	0.5	1.4×10^{-7}	2.6×10^{-9}	1.2×10^{-9}	53.9	112.2
1	0.1	1.4×10^{-6}	2.3×10^{-8}	4.3×10^{-8}	57.9	31.9
1	10	2.3×10^{-7}	4.1×10^{-9}	4.6×10^{-9}	55.9	49.6

As can be seen, the dependence on Γ is quite weak, and the quotient c_{lp}/c_{sp} is quite insensitive to the spectral index n as well. Let us for definiteness assume the canonical values $n = 1$ and $\Gamma = 0.5$ in what follows.

These values c_i would be the average values of the probability distributions for s_{sp} and s_{lp} if there where no experimental noise. We will now model the full probability distributions of the three experiments, including the contribution from experimental noise.

For a bulk flow experiment, the three components v_x, v_y and v_z of the velocity vector $\mathbf{v}$ are expected to be independent Gaussian random variables with zero mean, and

$$\langle|\mathbf{v}|^2\rangle = c_{lp}. \tag{8}$$

However, this is not quite the random variable s_{lp} that we measure, because of errors in distance estimation, *etc.* Denoting the difference between the observed and true bulk velocity vectors by ϵ, let us assume that the three components of ϵ are identically distributed and independent Gaussian random variables. This should be a good approximation, since even if the errors for individual galaxies are not, the errors in the average velocity ϵ will be approximately Gaussian by the Central Limit Theorem. Thus the velocity vector that we measure, $\mathbf{v} + \epsilon$, is also Gaussian, being the sum of two Gaussians. The variable that we actually measure is $s_{lp} = |\mathbf{v} + \epsilon|^2$, so

$$s_{lp} = \frac{1}{3}\left(c_{lp} + V_{lp}\right)\chi_3^2, \tag{9}$$

where χ_3^2 has a chi-squared distribution with three degrees of freedom, and V_{lp} is the variance due to experimental noise, *i.e.* the average variance that would be detected even if the true power spectrum were $P(k) = 0$. The fact that the expectation value of the detected signal s_{lp} (which is usually referred to as the *uncorrected* signal in the literature) exceeds the true signal c_{lp} is usually referred to as *error bias* (LP; Strauss Cen & Ostriker 1993 – hereafter SCO). Error bias is ubiquitous to all experiments of the type discussed in this paper, including CMB experiments, since the measured quantity is positive definite and the noise errors contribute squared. In the literature, experimentally detected signals are usually quoted after error bias has been corrected for, *i.e.* after the noise has been subtracted from the uncorrected signal in For LP, the uncorrected signal is 807 km/s, whereas the signal quoted after error bias correction is 689 km/s.

For the special case of the LP experiment, detailed probability distributions have been computed using Monte-Carlo simulations (LP, SCO), which incorporate such experiment-specific complications as sampling errors, asymmetry in the error ellipsoid, *etc.* To be used here, such simulations would need to be carried out for each value of c_{lp} under consideration. Since the purpose of this section is merely to give an example of the test formalism, the above-mentioned χ^2-approximation will be quite sufficient for our needs.

For the SP91 nine-point scan, the nine true values $\Delta T_i/T$ are expected to be Gaussian random variables that to a good approximation are independent. They have zero mean, and

$$\langle|\Delta T_i/T|^2\rangle = c_{sp}. \tag{10}$$

Denoting the difference between the actual and observed values by δ_i, we make the standard assumption that these nine quantities are identically distributed and independent Gaussian random variables. Thus the temperature fluctuation that we measure at each point, $\Delta T_i/T+\delta$, is again Gaussian, being the sum of two Gaussians. The variable that we actually measure is

$$s_{sp} = \frac{1}{9}\sum_1^9\left(\frac{\Delta T_i}{T}+\delta_i\right)^2 = \frac{1}{9}\left(c_{sp}+V_{sp}\right)\chi_9^2, \tag{11}$$

where χ_9^2 has a chi-squared distribution with nine degrees of freedom, and V_{sp} is the variance due to experimental noise, the error bias, *i.e.* the average variance that would be detected even if the true power spectrum were $P(k)=0$.

We will use only the signal from highest of the four frequency channels, which is the one likely to be the least affected by galactic contamination. Again, although Monte-Carlo simulations would be needed to obtain the exact probability distributions, we will use the simple χ^2-approximation here. In this case, the main experiment-specific complication is the reported gradient removal, which is a non-linear operation and thus does not simply lead to a χ^2-distribution with fewer degrees of freedom.

The amplitude of the COBE signal can be characterized by the variance in $\Delta T/T$ on an angular scale of $10°$. This number can be estimated from the COBE data set as $s_c = \sigma^2_{10°} = ((11.0\pm1.8)\times 10^{-5})^2$ (Smoot *et al.* 1992). The uncertainty in this quantity is purely due to instrument noise, and contains no allowance for cosmic variance. We must fold in the contribution due to cosmic variance in order to determine the probability distribution for s_c. We determined this probability distribution by performing Monte-Carlo simulations of the COBE experiment. We made simulated COBE maps with a variety of power spectra (including power laws with indices ranging from 0 to 3, as well as delta-function power spectra of the sort described in Section 4). We included instrumental noise in the maps, and excluded all points within $20°$ of the Galactic plane. By estimating s_c from each map, we were able to construct a probability distribution corresponding to each power spectrum. In all cases, the first three moments of the distribution were well approximated by

$$\begin{aligned}\mu_1 &\equiv \langle s_c\rangle = c_c,\\ \mu_2 &\equiv \langle s_c^2\rangle - \langle s_c\rangle^2 \le 0.063c_c^2 + 1.44\times10^{-21},\\ \mu_3 &\equiv \langle s_c^3\rangle = 0.009c_c^3.\end{aligned} \tag{12}$$

Furthermore, in all cases the probability distributions were well modeled by chi-squared distributions with the number of degrees of freedom, mean, and offset chosen to reproduce these three moments. Note that the magnitude of the cosmic variance depends on the shape of the power spectrum as well as its amplitude. The inequality in the above expression for μ_2 represents the largest cosmic variance of any of the power spectra we tested. Since we wish to set conservative limits on models, we will henceforth assume that the cosmic variance is given by this worst-case value. Thus we are assuming that the random variable $(s_c-s_0)/\Delta s$ has a chi-squared distribution with δ degrees of freedom, where

$$\begin{aligned}s_0 &= \mu_1 - 2\mu_2^2/\mu_3,\\ \Delta s &= \mu_3/4\mu_2,\\ \delta &= 8\mu_2^3/\mu_3^2.\end{aligned} \tag{13}$$

The results obtained using these three probability distributions are summarized in Tables 3abc. In 3a and 3b, $N=2$, and the question asked is whether LP is consistent with COBE and SP91, respectively. In Table 3c, $N=3$, and we test all three experiments for consistency simultaneously. In each case, the optimum normalization (proportional to the entries labeled "Signal") is different, chosen such that the consistency probability for the experiments under consideration is maximized. As can be seen, the last two tests rule out CDM at a significance level of 95%, *i.e.* predict that in an ensemble of universes, results as extreme as those we observe would be obtained less than 5% of the time. Note that using both COBE and SP91 to constrain LP yields a rejection that is no stronger than that obtained when ignoring COBE. In the latter case, the best fit is indeed that with no cosmological power at all, which agrees well with the observation of SCO that sampling variance would lead LP to detect a sizable bulk flow (before correcting for error bias) even if there where none.

Table 3a. Are LP and COBE consistent with CDM?

	LP	COBE	Combined
Noise	420 km/s	9.8 μK	
Signal	169 km/s	33.8 μK	
Noise+Signal	453 km/s	35.2 μK	
Detected	807 km/s	35.2 μK	
$\hat{\eta}$	0.046	1.00	0.046
$P(\eta < \hat{\eta})$	0.046	1.00	0.19

Table 3b. Are LP and SP91 consistent with CDM?

	LP	SP91	Combined
Noise	420 km/s	26.4 μK	
Signal	0 km/s	0 μK	
Noise+Signal	420 km/s	26.4 μK	
Detected	807 km/s	19.9 μK	
$\hat{\eta}$	0.023	0.35	0.0079
$P(\eta < \hat{\eta})$	0.023	0.35	0.046

Table 3c. Are LP, SP91 and COBE all consistent with CDM?

	LP	SP91	COBE	Combined
Noise	420 km/s	26.4 μK	9.8 μK	
Signal	168 km/s	26.9 μK	33.8 μK	
Noise+Signal	452 km/s	37.7 μK	35.1 μK	
Detected	807 km/s	19.9 μK	35.2 μK	
$\hat{\eta}$	0.046	0.039	0.97	0.0017
$P(\eta < \hat{\eta})$	0.046	0.039	0.97	0.046

Tables 3abc show the consistency probability calculations. The first line in each table gives the experimental noise, i.e. the detection that would be expected in the absence of any cosmological signal. The second line is the best-fit value for the cosmological signal c, the value that maximizes the combined consistency probability in the lower right corner of the table. The third line contains the expected value of an experimental detection, and is the sum in quadrature of the two preceding lines. The fourth line gives the goodness-of-fit parameter for each of the experiments, i.e. the probability that they would yield results at least as extreme as they did. The rightmost number is the combined goodness-of-fit parameter, which is the product of the others. The last line contains the consistency probabilities, the probabilities of obtaining goodness-of-fit parameters at least as low as those on the preceding line.

4. ALLOWING ARBITRARY POWER SPECTRA

In this section, we will derive the mathematical formalism for testing results from multiple experiments for consistency, without making any assumptions whatsoever about the power spectrum. This approach was pioneered by Juszkiewicz *et al.* (1987) for the case $N = 2$. Here we generalize the results to the case of arbitrary N. Despite the fact that the original optimization problem is infinite-dimensional, the necessary calculations will be seen to be of a numerically straightforward type, the case of N independent constraints leading to nothing more involved than numerically solving a system of N coupled non-linear equations. After showing this, we will discuss some inequalities that provide both a good approximation of the exact results and a useful qualitative understanding of them.

4.1. The Optimization Problem

Let us consider $N = n+1$ experiments numbered 0, 1, ..., n that probe the cosmological power spectrum $P(k)$. We will think of each experiment as measuring some weighted average of the power spectrum, and characterize an experiment E_i by its window function $W_i(k)$ as before.

Purely hypothetically, suppose we that we had repeated the same experiments in many different locations in the universe, and for all practical purposes knew the quantities $c_1, ..., c_n$ exactly. Then for which power spectrum $P(k)$ would c_0 be maximized, and what would this maximum be? If we experimentally determined c_0 to be larger than this maximum value, our results would be inconsistent, and we would be forced to conclude that something was fundamentally wrong either with our theory or with one of the experiments. In this section, we will solve this hypothetical problem. After this, it will be seen that the real problem, including cosmic variance and experimental noise, can be solved in almost exactly the same way.

The extremal power spectrum we are looking for is the solution to the following linear variational problem:

Maximize
$$\int_0^\infty P(k)W_0(k)dk \tag{14}$$
subject to the constraints that
$$\begin{cases} \int_0^\infty P(k)W_i(k)dk = c_i & \text{for } = 1, ..., n, \\ P(k) \geq 0 & \text{for all } k \geq 0. \end{cases} \tag{15}$$

This is the infinite-dimensional analogue of the so called linear programming problem, and its solution is quite analogous to the finite-dimensional case. In geometrical terms, we think of each power spectrum as a point in the infinite dimensional vector space of power spectra (tempered distributions on the positive real line, to be precise), and limit ourselves to the subset Ω of points where all the above constraints are satisfied. We have a linear function on this space, and we seek the point within the subset Ω where this function is maximized. We know that a differentiable functional on a bounded region takes its maximum either at an interior point, at which its gradient will vanish, or at a boundary point. In linear optimization problems like the one above, the gradient (here the variation with respect to P, which is simply the function W_0) is simply a constant, and will never vanish. Thus any maximum will always be attained at a boundary point. Moreover, from the theory of linear programming, we know that if there are n linear constraint equations, then the optimum point will be a point where all but at most n of the coordinates are zero. It is straightforward to generalize this result to our infinite-dimensional case, where each fixed k specifies a "coordinate" $P(k)$, and the result is that the solution to the variational problem is of the form
$$P(k) = \sum_{i=1}^{n} p_i \delta(k - k_i). \tag{16}$$
This reduces the optimization problem from an infinite-dimensional one to a $2n$-dimensional one, where only the constants p_i and k_i remain to be determined:

Maximize

$$\sum_{j=1}^{n} p_j W_0(k_j) \tag{17}$$

subject to the constraints that

$$\begin{cases} \sum_{j=1}^{n} p_j W_i(k_j) = c_i & \text{for } i = 1, ..., n, \\ p_i \geq 0 & \text{for } i = 1, ..., n. \end{cases} \tag{18}$$

This problem is readily solved using the method of Lagrange multipliers: defining the Lagrangian

$$L = \sum_{j=1}^{n} p_j W_0(k_j) - \sum_{i=1}^{n} \lambda_i \left[\sum_{j=1}^{n} p_j W_i(k_j) - c_i \right] \tag{19}$$

and requiring that all derivatives vanish leaves the following set of $3n$ equations to determine the $3n$ unknowns p_i, k_i and λ_i:

$$\begin{cases} W_0(k_i) - \sum_{j=1}^{n} \lambda_j W_j(k_i) = 0, \\ \left[W_0'(k_i) - \sum_{j=1}^{n} \lambda_j W_j'(k_i) \right] p_i = 0, \\ c_i - \sum_{j=1}^{n} p_j W_i(k_j) = 0. \end{cases} \tag{20}$$

Introducing matrix notation by defining the k_i-dependent quantities $A_{ij} \equiv W_j(k_i)$, $B_{ij} \equiv W_j'(k_i)$, $a_i \equiv W_0(k_i)$ and $b_i \equiv W_0'(k_i)$ brings out the structure of these equations more clearly: If $p_i \neq 0$, then

$$\begin{cases} A\lambda = \mathbf{a}, \\ B\lambda = \mathbf{b}, \\ A^T \mathbf{p} = \mathbf{c}. \end{cases} \tag{21}$$

If A and B are invertible, then eliminating λ from the first two equations yields the following system of n equations to be solved for the n unknowns k_1, ..., k_n:

$$A^{-1}\mathbf{a} = B^{-1}\mathbf{b}. \tag{22}$$

Although this system is typically coupled and non-linear and out of reach of analytical solutions for realistic window functions, solving it numerically is quite straightforward. A useful feature is that once this system is solved, $\mathbf{a}$, $\mathbf{b}$, A and B are mere constants, and the other unknowns are simply given by matrix inversion:

$$\begin{cases} \lambda = A^{-1}\mathbf{a} \\ \mathbf{p} = \left(A^{-1}\right)^T \mathbf{c} \end{cases} \tag{23}$$

Since the non-linear system (22) may have more than one solution, all solutions should be substituted back into (17) to determine which one is the global maximum. Furthermore, to make statements about the solution to our original optimization problem (14), we need to consider also the case where one or more of the n variables p_1,...,p_n vanish. If exactly m of them are non-vanishing, then without loss of generality, we may assume that these are the first m of the n variables. Thus we need to solve the maximization problem (17) separately for the cases where $P(k)$ is composed of n delta functions, $n-1$ delta functions, *etc.*, all the way down to the case where $P(k)$ is single delta function. These solutions should then be substituted back into (14) to determine which is the global maximum sought in our original problem. Thus the solutions depend on the window functions W_i and the signals c_i in the following way:

- From the window functions alone, we can determine a discrete and usually finite number of candidate wavenumbers k where delta functions can be placed.
- The actual signals c_i enter only in determining the coefficients of the delta functions in the sum, i.e. in determining what amount of power should be hidden at the various candidate wavenumbers.

If we have found an the optimal solution, then a small change in the signal vector $\mathbf{c}$ will typically result in a small change in $\mathbf{p}$ and no change at all in the number of delta functions in $P(k)$ or their location. If $\mathbf{c}$ is changed by a large enough amount, the delta functions may suddenly jump and/or change in number as a different solution of (17) takes over as global optimum or one of the coefficients p_i becomes negative, the latter causing the local optimum to be rejected for constraint violation. Thus within certain limits, we get the extremely simple result that for the optimal power spectrum $P(k)$,

$$c_0 = \int_0^\infty P(k)W_0(k)dk = \left(A^{-1}\mathbf{a}\right)\cdot\mathbf{c}. \tag{24}$$

Thus within these limits, c_0 depends linearly on the observed signal strengths c_i. This is exactly analogous to what happens in linear programming problems.

4.2. A Useful Inequality

Before proceeding further, we will attempt to provide a more intuitive understanding of the results of the previous section, and show how to determine how complicated a calculation is justified. For the special case of only a single constraint, *i.e.* $n=1$, we obtain simply $P(k) = p_1\delta(k-k_1)$, where k_1 is given by

$$W_0'(k_1)W_1(k_1) = W_1'(k_1)W_0(k_1). \tag{25}$$

For the case of n constraints, let us define the functions

$$f_i \equiv \frac{W_0(k)}{W_1(k)}c_i.$$

Then we see that for $n=1$, k_1 is simply the wavenumber for which the function f_1 is maximized, and that the maximum signal possible is simply $c_0 = f_1(k_1)$. Thus the maximum signal in experiment 0 that is consistent with the constraint from experiment i is obtained when the power is concentrated where the function f_i is large. In other words, if we want to explain a high signal c_0 in the face of low signals in several constraining experiments, then the best place to hide the necessary power from the i^{th} experiment is where f_i takes its maximum. These functions are plotted in Figure 4 for the experiments discussed in Section 3, the optimization problem being the search for the maximum LP signal that is consistent with the constraints from SP91 and COBE. For illustrative purposes, we here assume that c_{sp} and c_c are known exactly, and given by the detected signals $\hat{s}_{sp}^{1/2} \approx 19.9\,\mu$K and $\hat{s}_c^{1/2} \approx 33.8\,\mu$K (we will give a proper treatment of cosmic variance and noise in the following section). Using the $n=1$ constraint for each constraining experiment separately, the smallest of the functions thus sets an upper limit to the allowed signal $c_0 = c_{sp}$. Thus the limit is given by the highest point in the hatched region in Figure 4, *i.e.*

$$c_0 \le c_{max}^{(1)} \equiv \sup_k \min_i f_i(k). \tag{26}$$

We see that using the SP91 constraint alone, the LP signal would be maximized if all power were at $k \approx (940\mathrm{Mpc})^{-1}$. Since this flagrantly violates the COBE constraint, the best place to hide the power is instead at $k \approx (100\mathrm{Mpc})^{-1}$.

By using the above formalism to impose all the constraints at once, the allowed signal obviously becomes lower. If the constraints are equalities rather than inequalities, then this stronger limit can never lie below the value at $(k_* \approx 250\text{Mpc})^{-1}$, where $f_{sp}(k_*) = f_c(k_*)$, since this is the signal that would result from a power spectrum of the form $P(k) \propto \delta(k - k_*)$. Thus for the particular window functions in our example, where the constraint from the $n = 2$ calculation cannot be more than a factor $f_{sp}(80\text{Mpc})/f_{sp}(250\text{Mpc}) \approx 1.05$ stronger than the simple $n = 1$ limits, the latter are so close to the true optimum that they are quite sufficient for our purposes. If the constraints are upper limits rather than equalities, then the limit on c_0 is more relaxed, and is always the uppermost point in the hatched region, *i.e.* $c^{(1)}_{max}$.

4.3. Including Noise and Cosmic Variance

To correctly handle cosmic variance and instrumental noise, we need to use the formalism developed in Section 2. Thus given the probability distributions for the various experimental results s_i, we wish to find the power spectrum for which the consistency probability η is maximized. This optimization problem, in which all experiments are treated on an equal footing, will be seen to lead directly to the asymmetric case above where the signal in one is maximized given constraints from the others. For definiteness, we will continue using the example with the LP, SP91 and COBE experiments. As seen in Section 2, the source of the low consistency probabilities is that $\hat{s}_{lp}$ is quite high when compared to $\hat{s}_{sp}$ and $\hat{s}_c$. Thus it is fairly obvious that for the power spectrum that maximizes the consistency probability, we will have $\hat{s}_{lp} > \langle s_{lp} \rangle$, whereas $\hat{s}_{sp} < \langle s_{sp} \rangle$ and $\hat{s}_c < \langle s_c \rangle$, so we can neglect power spectra that do not have this property. Let us first restrict ourselves to the subset of these power spectra for which $c_{lp} = D$ and $c_c = E$, where D and E are some constants. Then these power spectra all predict the same probability distributions for s_{lp} and s_c. The consistency probability η is clearly maximized by the power spectrum that maximizes $\langle s_{sp} \rangle$, and this will be a linear combination of one or two delta functions as shown in Section 4.1. The key point is that since the locations of these delta functions are independent of D and E (within the range discussed in 4.1), the infinite-dimensional optimization problem reduces to the following two simple steps:

1. Solve for the optimal number of delta functions m and their locations k_i as described in section 4.1
2. Find the m coefficients p_i for which the power spectrum $P(k) = \sum_{i=1}^{m} p_i \delta(k - k_i)$ maximizes the consistency probability.

4.4. Power spectrum independent constraints on LP, SP91 and COBE

When applying the above consistency test to the LP, SP91 and COBE experiments, we obtain exactly the same consistency probability as in Table 3b. The reason for this is that the optimal normalization turns out to be zero. This will obviously change if the LP error bars become smaller in the future. Thus dropping the CDM assumption does not improve the situation at all, which indicates that main source of the inconsistency must be something other than the CDM model.

In anticipation of future developments, consistency probabilities were also computed for a number of cases with less noise in the LP experiment. Comparing only LP and SP91, the optimum power spectrum has a delta function at $k \approx (941\text{Mpc})^{-1}$. When including all three experiments, treating the COBE and SP91 constraints as upper limits, the optimum power spectrum has a single delta function at $k \approx (79\text{Mpc})^{-1}$, so the addition of COBE strengthens the constraint only slightly, due to the flatness of f_{sp} in Figure 4. Interestingly, for all these cases with smaller LP error bars, consistency probabilities were found to be almost as low when allowing arbitrary power spectra as for the CDM case. This is again attributable to the flatness of f_{sp}, since weighted averages of a flat function are fairly independent of the shape of the weight function (here the power spectrum).

5. DISCUSSION

We have developed a formalism for testing multiple cosmological experiments for consistency. As an example of an application, we have used it to place constraints on bulk flows of galaxies using the COBE and SP91 measurements of fluctuations in the cosmic microwave background. It was found that taken at face value, the recent detection by Lauer and Postman of a bulk flow of 689 km/s on scales of $150h^{-1}$Mpc is inconsistent with SP91 within the framework of a CDM model, at a significance level of about 95%. However, interestingly, this cannot be due solely to the CDM assumption, since the LP result was shown to be inconsistent with COBE and SP91 at the same significance level even when no assumptions whatsoever were made about the power spectrum. This leaves four possibilities:

1. The window functions are not accurate.
2. Something is wrong with the quoted signals or error bars for at least one of the experiments,
3. The observed fluctuations cannot be explained within the framework of gravitational instability and the Sachs-Wolfe effect.
4. The random fields are not Gaussian,

Case 1 could be attributed to a number of effects: If $\Omega \neq 1$, then both the calculation of the Sachs-Wolfe effect (which determines W_{sp} and W_{cobe}) and the growth of velocity perturbations (which determines W_{lp}) are altered. If the universe became reionized early enough to rescatter a significant fraction of all CMB photons, then small scale CMB anisotropies were suppressed, which would lower W_{sp}. A quantitative treatment of these two cases will be given in a future paper. Other possible causes of 1 include a significant fraction of the density perturbations being isocurvature (entropy) perturbations or tensor-mode perturbations (gravity waves). Apart from these uncertainties, we have made several simplifying assumptions about the window functions for LP and SP91. To obtain more accurate consistency probabilities than those derived in the present paper, a more accurate LP window function should be used that incorporates the discreteness and the asymmetry of the sample of Abell clusters used. This can either be done analytically (Feldman & Watkins 1993) or circumvented altogether by performing Monte-Carlo simulations like those of LP or SCO, but for the whole family of power spectra under consideration.

As to case 2, there has been considerable debate about both the LP and the SP91 experiments. A recent Monte-Carlo Simulation of LP by SCO basically confirms the large error bars quoted by LP. As is evident from the flatness of the LP curve in Figure 3, it will be impossible to make very strong statements about inconsistency until future experiments produce smaller error bars. With the SP91 experiment, a source of concern is the validity of using only the highest of the four frequency channels to place limits, even though it is fairly clear that the other three channels suffer from problems with galactic contamination. The situation is made more disturbing by the fact that a measurement by the balloon-borne MAX experiment (Gundersen *et al.* 1993) has produced detections of degree-scale fluctuations that that are higher than those seen by SP91, and also higher than another MAX measurement (Meinhold *et al.* 1993). Other recent experiments that have detected greater fluctuations include ARGO (de Bernardis *et al.* 1993), PYTHON (Dragovan *et al.* 1993) and MSAM (Cheng *et al.* 1993). On the other hand, SP91 has been used only as an upper limit in our treatment, by including only the Sachs-Wolfe effect and neglecting both Doppler contributions from peculiar motions of the surface of last scattering and intrinsic density fluctuations at the recombination epoch. If these effects (which unfortunately depend strongly on parameters such as h and Ω_b) where included, the resulting constraints would be stronger.

Case 3 might be expected if the universe underwent a late-time phase transition, since this could generate new large-scale fluctuations in an entirely non-gravitational manner.

In the light of the many caveats in categories 1 and 2, the apparent inconsistency between LP and SP91 (Jaffe *et al.* 1993) is hardly a source of major concern at the present time, and it does not appear necessary to invoke 3 or 4. However, we expect the testing formalism developed in this paper to be able to provide many useful constraints in the future, as more experimental data is accumulated and error bars become smaller.

The authors wish to thank Michael Strauss, Bhuvnesh Jain, Douglas Scott, Martin White, Joseph Silk and an anonymous referee for many useful comments and suggestions.

REFERENCES

Alcock, C, Akerlof, C. W., Allsman, R. A., Axelrod, T. S. and others 1993, Nature, 365, 621.

de Bernardis *et al.* 1993, Ap. J. Lett., in press.

Bond, J.R. & Efstathiou, G. 1987, MNRAS, 226, 655.

Bond, J.R., Efstathiou, G., Lubin, P.M., & Meinhold, P.R. 1991, Phys. Rev. Lett., 45, 1980.

Cen, R., Gnedin, N. Y., Koffmann, L. A., & Ostriker, J. P. 1992, Ap. J. Lett., 399, L11.

Cen, R., Ostriker, J. P. & Peebles, P. J. E. 1993, Ap. J., 415, 423.

Cheng, E. *et al.* 1993, Ap. J. Lett., in press.

Dodelson, S. & Jubas, J.M. 1993, Phys. Rev. Lett., 70, 2224.

Dragovan, M. *et al.* 1993, preprint.

Feldman, H. A. & Watkins, R. 1993, preprint.

Feynman, R. P. 1939, Phys. Rev., 56, 340.

Gaier, T., Schuster, J., Gundersen, J. O., Koch, T., Meinhold, P. R., Seiffert, M. & Lubin, P. M., 1992, ApJ, 398, L1.

Gnedin, N. Y. & Ostriker, J. P.1992, Ap. J., 400, 1.

Górski, K.1991, Ap. J. Lett., 370, L5.

Górski, K.1992, Ap. J. Lett., 398, L5.

Gundersen, J.O., Clapp, A.C., Devlin, M., Holmes, W., Fischer, M.L., Meinhold, P.R., Lange, A.E., Lubin, P. M., Richards, P.L., & Smoot, G.F. 1993, Ap. J. Lett., 413, L1.

Hancock *et al.* 1994, Nature, 367, 333.

Jaffe, A., Stebbins, A., & Frieman, J.A 1993. Ap. J. (in press)

Juszkiewicz, R. 1993, Private communication.

Juszkiewicz, R., Górski, K., & Silk, J.1987, Ap. J. Lett., 323, L1.

Kamionkowski, M. & Spergel, D. 1993, preprint.

Kolb, E. W. & Turner, M.S. 1990, *The Early Universe*, Addison Wesley

Lauer, T. & Postman, M. 1993, preprint. (LP)

Maddox, S. J., Efstathiou, G., Sutherland, W. J. & Loveday, J. 1990, MNRAS, 242, 43.

Meinhold, P. R. & Lubin, P. M. 1991, Ap. J. Lett., 370, L11.

Meinhold, P. R., Clapp, A.C., Cottingham, D., Devlin, M., Fischer, M.L., Gundersen, J.O., Holmes, W., Lange, A.E., Lubin, P.M., Richards, P.L., & Smoot, G.F. 1993, Ap. J. Lett., 409, L1.

Oukbir, J. & Blanchard, A. 1992, Astr. Ap., 262, L21.

Padmanabhan, T. 1993, *Structure Formation in the Universe*, Cambridge Univ. Press, New York

Peebles, P. J. E. 1984, Ap. J., 284, 439.

Peebles, P. J. E. 1987, Ap. J., 315, L73.

Schlegel, D., Davis, M., Summers, F. & Holtzman, J. 1993, preprint.

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

Strauss, M., Cen, R. & Ostriker, J.P. 1993, preprint. (SCO)

Suto, Y., Górski, K. Juszkiewicz, R., & Silk, J.1988, Nature, 332, 328.

Watson, R.A. & Gutiérrez de la Cruz, C.M. 1993, Ap. J. Lett., 419, L5.

White, M., Krauss, L., & Silk, J. 1993, Ap. J., 418, 535.

APPENDIX A: COMPARISON OF GOODNESS-OF-FIT PARAMETERS

In this appendix, we compare the performances of the probability product and the likelihood product as goodness-of-fit parameters.

First of all, what do we mean by a goodness-of-fit parameter η being good? That η leads to the correct model being ruled out at some confidence level x cannot be held against it — by definition of significance level, this happens a fraction $(1-x)$ of the time. Rather, the conventional criterion for rating goodness-of-fit parameters is *rejection power*: given a model and a set of observations, one η is said to be more powerful than another if it rejects the model at a higher level of significance. An example of a very stupid goodness-of-fit parameter, which in fact has the lowest rejection power possible, is a random variable η drawn from a uniform distribution on $[0,1]$, thus containing no information whatsoever about the model or the observed data. Use of this parameter will reject the model at 95% confidence only 5% of the time, even if the data is blatantly inconsistent with the model.

Both η_l and η_p can be "fooled"...

Given a random variable s_i with probability distribution f_i, we define the corresponding random variable for likelihood by $L_i = f_i(s_i)/f_{max}$, where we chose the normalization constant $f_{max} \equiv \max_x f_i(x)$ so that we will always have $0 \le L_i \le 1$. Thus for N experiments, the likelihood product

$$\eta_l \equiv \prod_{i=1}^{N} L_i$$

will also be a random variable on the interval $[0,1]$.

It is easy to construct examples where either the probability product η_p (as defined in Section 2.1) or the likelihood product η_l give very low rejection power. The Achilles' heel of the probability product is multimodal distributions, where values near the mean are rather unlikely. For example, suppose $N=1$ and we have a double-humped distribution such as

$$f_1(s) \propto s^2 e^{-s^2}.$$

If we observe a value $\hat{s}_1 = 0$, then η_l would reject the model with 100% confidence whereas the probability product fails miserably, rejecting with 0% confidence since $\hat{s}_1$ equals the mean. The likelihood product, on the other hand, has the weakness that the highest likelihood may be attained far out in the tail of the distribution. Suppose for instance that $N=5$ and we have triangle distributions

$$f_i(s) = 2s$$

on the interval $0 \le s \le 1$. If all five observed values $\hat{s}_i$ lie between 0.999 and 1, something is clearly wrong with the model. Since $\hat{\eta}_p > 0.004^5$, using Eq. (3) indeed rejects the model at a confidence level exceeding $1 - F_\eta(0.004^5) \approx 99.999997\%$. Yet the likelihood product is near its maximum value: $\hat{\eta}_l \ge 0.999^5 \approx 0.995$, which gives virtually no rejection power at all.

...but they usually give similar results.

It is important to note that neither of the two examples above were particularly physical. The random variables arising from cosmic variance have Gaussian or chi-squared distributions, and the same tends to holds for the various experimental noise distributions. Thus the probability distributions to which our goodness-of-fit parameter is applied in this paper are unimodal (which eliminates the first example) and taper off to zero smoothly (which eliminates the second example). Hence for cosmological applications, goodness-of-fit parameters should not be rated by their performance with such pathological distributions, but rather by their rejection power when applied to continuous, unimodal distributions. We will now compare the performance of η_p and η_l for a few such cases.

For a symmetric *exponential* distribution

$$f_i(s) = \frac{1}{2}e^{-|s|},$$

it is easy to see that the likelihood L_i has a uniform distribution. This means that η_p and η_l will have identical distributions, for arbitrary N. It is straightforward to show that the same holds for symmetric *triangle* distributions

$$f_i(s) = 1 - |s|.$$

A third case where η_p and η_l give identical results is when $N = 1$ and f is any smooth unimodal function.

For a *Gaussian* distribution

$$f_i(s) = (2\pi)^{-1/2}e^{-s^2/2},$$

we have

$$P(L_i < x) = 2\mathrm{erfc}\left[(-2\ln x)^{1/2}\right].$$

Although the probability distribution of η_l appears not to be expressible in terms of elementary functions for arbitrary n, it is easy to show that η_l has a uniform distribution for the special case $N = 2$. Thus the likelihood product gives rejection at a confidence level of

$$P(\eta_l < \hat{\eta}_l) = e^{-(\hat{s}_1^2+\hat{s}_2^2)/2}.$$

Comparing this with the corresponding confidence level based on η_p shows a remarkable agreement between the two methods. Within the disc $\hat{s}_1^2 + \hat{s}_2^2 < 4$, over which $P(\eta < \hat{\eta})$ varies with many orders of magnitude, the two methods never differ by more than a factor of two. Thus, at worst, one may yield a confidence level of say 99.98% where the other yields 99.99%. The probability product is stronger in slightly more than half of the $(\hat{s}_1, \hat{s}_2)$-plane, roughly for regions that are more than 20° away from any of the coordinate axes.

In conclusion, we have seen that for unimodal, continuous probability distributions, the likelihood product and the probability product tend to give fairly similar — in a few special cases even identical — results. Thus chosing one parameter over the other is more a matter of personal preference than something that is likely to seriously affect any scientific conclusions. There is however one important practical consideration: The probability distribution of η_l depends on the probability distributions of the random variables s_i. This means that, apart from a few fortuitous special cases such as described above, it can generally not be calculated analytically. Rather, it must be computed numerically, through numerical convolution or Monte Carlo simulation. The probability distribution for η_p, on the other hand, is always known analytically, as given by Eq. (3), so the probability product is considerably simpler to use.

APPENDIX B: WINDOW FUNCTIONS

The results of CMB anisotropy experiments can be conveniently described by expanding the temperature fluctuation in spherical harmonics:

$$\frac{\Delta T}{T}(\hat{\mathbf{r}}) = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\hat{\mathbf{r}}). \tag{27}$$

(The monopole and dipole anisotropies have been removed from the above expression, since they are unmeasurable.) If the fluctuations are Gaussian, then each coefficient a_{lm} is an independent Gaussian random variable with zero mean (Bond and Efstathiou 1987). The statistical properties of the fluctuations are then completely specified by the variances of these quantities

$$C_l \equiv \left\langle |a_{lm}|^2 \right\rangle. \tag{28}$$

(The fact that the variances are independent of m is an immediate consequence of spherical symmetry.) Different CMB experiments are sensitive to different linear combinations of the C_l's:

$$S = \sum_{l=2}^{\infty} F_l C_l, \tag{29}$$

where S is the ensemble-averaged mean-square signal in a particular experiment, and the "filter function" F_l specifies the sensitivity of the experiment on different angular scales. The filter functions for COBE and SP91 are

$$\begin{aligned} F_l^{(cobe)} &= \frac{(2l+1)}{4\pi} e^{-\sigma_c^2 (l+\frac{1}{2})^2}, \\ F_l^{(sp)} &= 4 e^{-\sigma_s^2 (l+\frac{1}{2})^2} \sum_{m=-l}^{l} H_0^2(\alpha m), \end{aligned} \tag{30}$$

where H_0 is a Struve function. $\sigma_c = 4.25°$ and $\sigma_s = 0.70°$ are the r.m.s. beamwidths for the two experiments, and $\alpha = 1.5°$ is the amplitude of the beam chop (Bond *et al.* 1991, Dodelson and Jubas 1993, White *et al.* 1993).

For Sachs-Wolfe fluctuations in a spatially flat Universe with the standard ionization history, the angular power spectrum C_l is related to the power spectrum of the matter fluctuations in the following way (Peebles 1984, Bond and Efstathiou 1987):

$$C_l = \frac{8}{\pi \tau_0^4} \int_0^{\infty} dk P(k) \bar{j}_l^2(k). \tag{31}$$

Here τ_0 is the conformal time at the present epoch, and

$$\bar{j}_l(k) \equiv \int j_l(k\tau) V(\tau)\, d\tau, \tag{32}$$

where j_l is a spherical Bessel function. The visibility function V is the probability distribution for the conformal time at which a random CMB photon was last scattered. $\bar{j}_l(k)$ is therefore the average of $j_l(k\tau)$ over the last scattering surface. We have used the V of Padmanabhan (1993).

We can combine equations (29), (30), and (31) to get the window functions for the two experiments:

$$\begin{aligned} W_{cobe} &= \frac{2}{\pi^2 k^2 \tau_0^4} \sum_{l=2}^{\infty} \bar{j}_l^2(k) e^{-\sigma_c^2 \left(l+\frac{1}{2}\right)^2} (2l+1) \\ W_{sp} &= \frac{32}{\pi k^2 \tau_0^4} \sum_{l=2}^{\infty} \bar{j}_l^2(k) e^{-\sigma_s^2 \left(l+\frac{1}{2}\right)^2} \sum_{m=-l}^{l} H_0^2(\alpha m) \end{aligned} \tag{33}$$

The mean-square bulk flow inside of a sphere of radius a is (see, *e.g.*, Kolb and Turner 1990)

$$\langle v^2 \rangle = \int dk P(k) \frac{18}{\pi^2 \tau_0^2} \frac{j_1^2(ka)}{(ka)^2}. \tag{34}$$

However, we must make two corrections to this result before applying it to the LP data. This formula applies to a measurement of the bulk flow within a sphere with an infinitely sharp boundary. In reality, errors in measuring distances cause the boundary of the spherical region to be somewhat fuzzy. If we assume that distance measurements are subject to a fractional error ϵ, then the window function must be multiplied by $e^{-(\epsilon k a)^2}$. We have taken $\epsilon = 0.16$, the average value quoted by LP. It should be noted that this value varies from galaxy to galaxy in the LP sample, due to the distance estimation technique used, and that a more accurate window function that reflects the discrete locations of the Abell clusters used in the survey should take this into account.

The second correction has to do with the behavior of the window function at small k. Equation (34) applies to the velocity relative to the rest frame of the Universe. The velocity measured by LP is with respect to the CMB rest frame. If there is an intrinsic CMB dipole anisotropy, then these two reference frames differ. Therefore, we must include in equation (29), a term corresponding to the intrinsic CMB dipole. This correction was first noticed by Górski (1991). After applying both of these corrections, the LP window function is

$$W_{lp} = \frac{18}{\pi^2 \tau_0^2} \left(\frac{j_1(ka)}{ka} e^{-(\epsilon k a)^2} - \frac{\bar{j}_1(k\tau_0)}{k\tau_0} \right)^2. \tag{35}$$

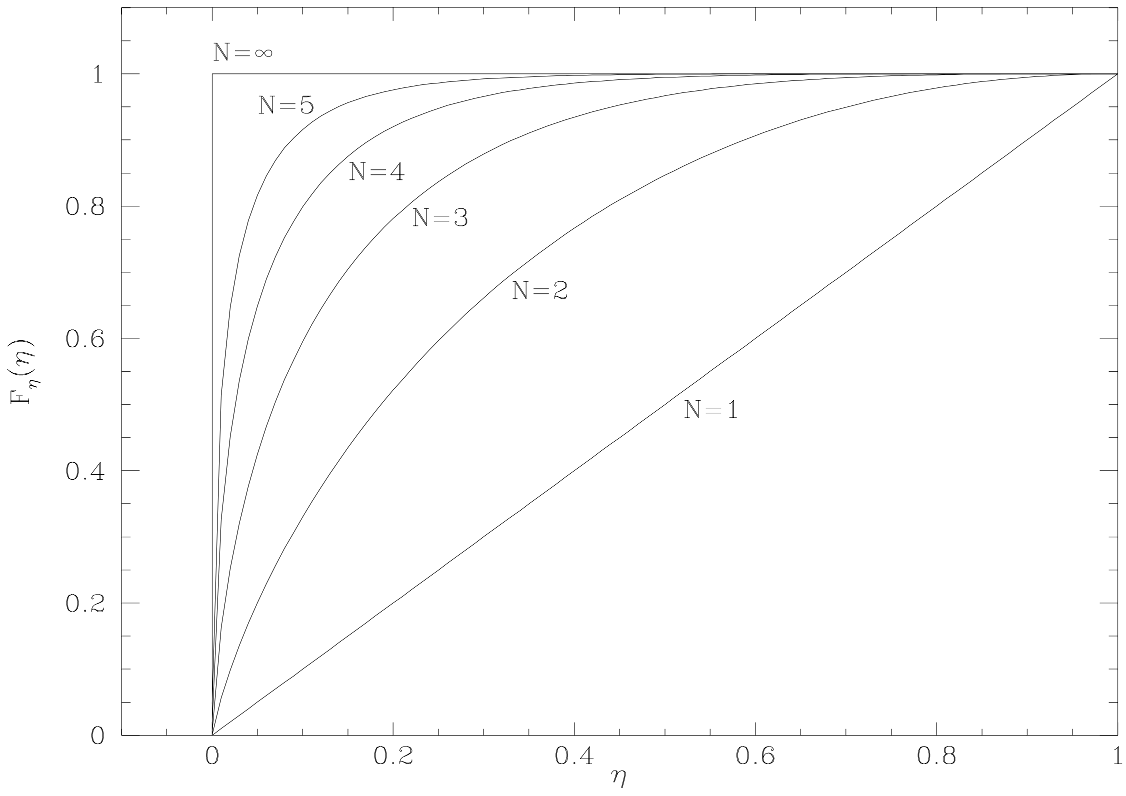

Figure 1. The function F_η.
The cumulative probability distribution for the goodness-of-fit parameter η is plotted for a few different n-values.

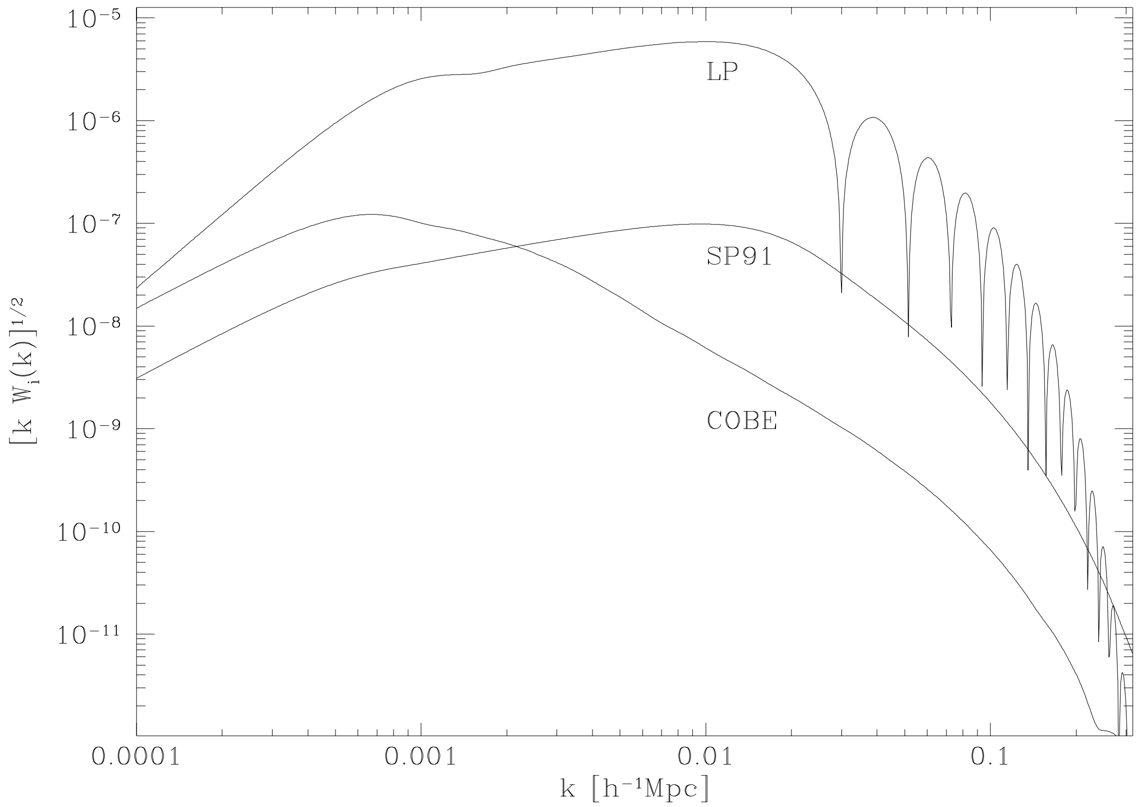

Figure 2. Window Functions.
The window functions of the Lauer-Postman bulk flow measurement (LP), the South Pole 1991 nine point scan (SP91), and the COBE DMR 10° *pixel r.m.s. measurement (COBE) are plotted as a function of comoving wavenumber* k*.*

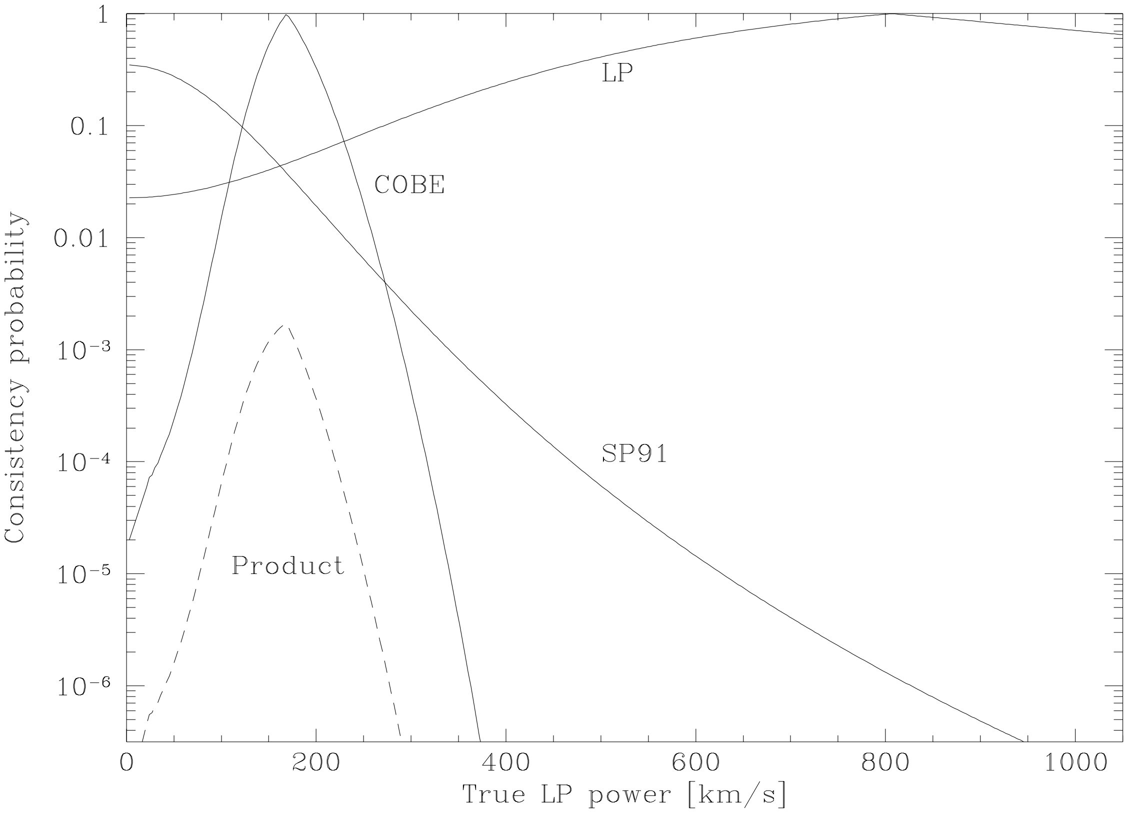

Figure 3. Consistency Probabilities
The probability that the the Lauer-Postman bulk flow measurement (LP), the COBE DMR experiment and the South Pole 1991 nine point scan (SP91) are consistent with CDM is plotted as a function of the normalization of the power spectrum. The normalization is expressed in terms of the expected bulk flow in a LP measurement. The dashed line is the product of these three probabilities, and takes a maximum for a normalization corresponding to 168 km/s.

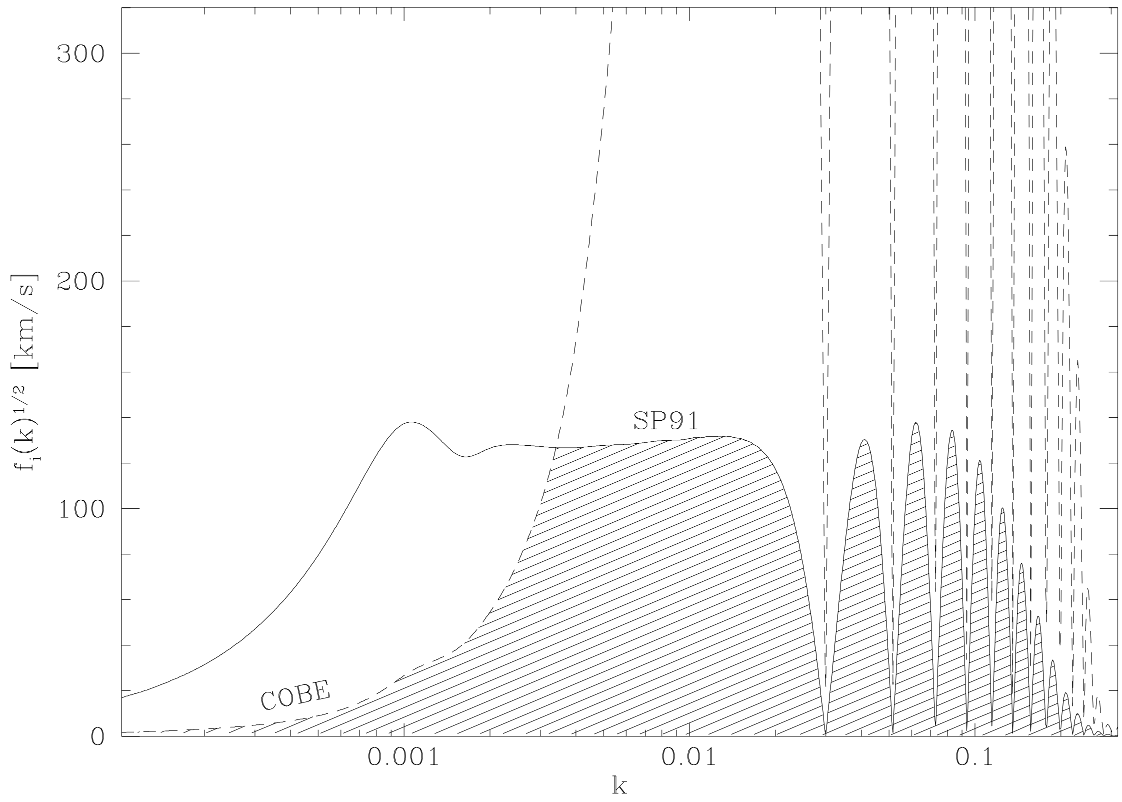

Figure 4. The best places to hide power

The functions f_{sp} (solid line) and $fcobe$ (dashed line) are plotted as a function of wavenumber k. The shaded region, i.e. the area lying beneath both curves, constitutes the LP bulk flows that would be consistent with the SP91 and COBE experiments using the $N = 1$ constraints only, when the power spectrum is a single delta function located at k.