Statistical Analysis of Cosmic Microwave Background Anisotropy

by

Emory Freeman Bunn

A.B. (Princeton University) 1989
M.A. (University of California at Berkeley) 1993

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Physics

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Joseph Silk, Chair

Professor Marc Davis

Professor Philip Stark

1995

The dissertation of Emory Freeman Bunn is approved:

Chair Date

Date

Date

University of California, Berkeley
1995

Statistical Analysis of
Cosmic Microwave Background Anisotropy

Contents

List of Figures

List of Tables

Acknowledgments

This document, like all of the research I have done at Berkeley, would never have come into being without the support and collaboration of many people. Chief among them is my adviser, Joe Silk. Joe's prodigious creativity and knowledge of astrophysics (knowledge that is as remarkable for its breadth as for its depth) have been invaluable to me. His patience, support, and encouragement have been unflagging. He never gets mad at me, even when I deserve it.

The work presented in this dissertation is the result of a fairly large and ever-shifting collaboration. Various portions were produced with the help of Yehuda Hoffman, Douglas Scott, Naoshi Sugiyama, Max Tegmark, and Martin White. All of these people, but especially Martin and Douglas, have made great contributions to my education as a physicist.

Chapter 5 in particular bears the very strong imprint of various collaborators. I have never solved the Boltzmann equation in my life: all of the angular power spectra in this chapter (and indeed in the entire dissertation) are the work of Martin and Naoshi. Furthermore, Martin and Douglas computed all of the small-scale normalizations in Tables 5.1, 5.2, and 5.3, and Martin made many of the figures in Section 5.3.

Berkeley swarms with clever astrophysicists, and I have been fortunate to learn from many of them. David Schlegel in particular has helped me in more ways than I can recall. From the moment I began to study astrophysics, he has been one of my first resources for astronomical information. He has helped me with countless computer problems (including deciphering the COBE data format, without which help this dissertation would be a lot shorter). He convinced me to start biking, and he slows down for me when we go uphill. But most importantly, he has been a very close and valued friend for many years.

On countless occasions, Wayne Hu has helped me along with patient and clear explanations of cosmological arcana; he has also been an ideal officemate and roommate at various conferences, and a worthy adversary in the thesis race. Max Tegmark (*né* Shapiro) and I

have struggled together with several problems in statistics, linear algebra, and cosmology; he is also an excellent hiking companion. Other people who have been instrumental in my cosmological education include Marc Davis, Karl Fisher, Ofer Lahav, Charley Lineweaver, Paul Richards, Mark Srednicki, and Saleem Zaroubi. Marc also provided insightful criticism of this dissertation, as did Joe Silk and Eric Gawiser.

Philip Stark has set me straight on statistical matters on numerous occasions, and has heroically suppressed his anti-Bayesian fervor in our conversations. His thorough reading of this dissertation resulted in dramatic improvements in its prose style and clarity.

It is a pleasure to acknowledge the support of my close friends. Tim Wagner has been the best roommate and friend one could hope for, and not simply because of his prodigious Calphalon collection, or because he set up my window manager. He is a man of remarkable warmth, wit, and generosity, and I count myself lucky to have lived with him. Ashley Silverburg has created many happy memories for me and has watched out for me during difficult times; knowing her has enriched my life. Among the many other people whose friendship and support have been of great value to me are Nasreen Chopra, Matt Craig, Chris Genovese, Sabrina Grannan, David Handin, Mike Hase, and Alison Reeve.

It is a pleasure to thank Anne Takizawa and Donna Sakima, who, with consummate valor and skill, shield all physics graduate students from the onslaught of administrators; and Lexi Moustakas and Dan Stevens, who managed to tolerate me as an officemate.

Finally, and most importantly, I want to thank my family. Unfortunately, acknowledgments are nearly always written in superlatives, leaving no words adequate to describe those people for whom superlatives are truly appropriate. Mom, Dad, George, and Andy, you really are the best. Your boundless love and support have meant more to me than you can imagine.

Chapter 1

Cosmology and the Cosmic Microwave Background

> A man said to the universe:
> "Sir I exist!"
> "However," replied the universe,
> "The fact has not created in me
> A sense of obligation."
> Stephen Crane

1.1 Introduction

In 1899, when Crane painted this bleak picture of a Universe indifferent to human concerns [43], humanity had no idea just how vast and empty space is. Although the Earth had long since been dethroned from its location at the center of the Universe [39], we were still unaware of the true enormity of creation and of our thoroughly unremarkable location in it. Over the following decades we would discover that our Sun is located in an undistinguished spot in the outskirts of the Milky Way Galaxy and that our galaxy is a typical spiral with nothing to set it apart from the billions of others sprinkled more or less uniformly through space. In comparison with our current view, the Universe of Crane's era was downright cozy!

We therefore have even more reason today than Crane did to observe that the Universe is under no obligation to its denizens; in particular, it is under no obligation to make itself comprehensible to us. However, for as long as there have been scientists, there have been people willing to make the optimistic assumptions that the Universe is comprehensible and that by diligent study and observation we will be able to understand and predict its behavior. Cosmologists are perhaps the most guilty of this unfounded optimism; they hope to understand not just a piece of the Universe, but the whole thing. Over the past several decades, this optimism has been at least partially vindicated: we appear to have discovered at least the rudiments of a genuine, quantitative, predictive theory of physical cosmology.

Testing this theory has not been easy. It has been observed that "to a first approximation, all observables in cosmology are impossible to measure" [108]. This statement is not as much of an exaggeration as we might wish: even the most fundamental cosmological observables such as the Hubble parameter H_0 are known only to a rough approximation [95]. Indeed, as recently as 1963, Peter Scheuer observed that "there are only $2\,{}^1\!/_2$ facts in cosmology" [119]. Despite these difficulties, our knowledge of cosmology has made enormous strides since Scheuer gave this pessimistic assessment. We have accumulated a substantial body of evidence in support of the standard "hot big bang model," and we have good reason to believe that this model accurately describes many aspects of the evolution of the Universe from a very early epoch (say about 10^{-6} s after the Big Bang) to the present.

In brief, the big bang model proposes that the early Universe was extremely hot and dense and that it expanded and cooled to its present state over a period of about ten to twenty billion years. The Universe is presumed to be approximately homogeneous and isotropic (when averaged over suitably large scales). Given these assumptions, we can use general relativity to determine the spacetime geometry of the Universe, which turns out to be one of a simple family of models, the Friedmann-Robertson-Walker (FRW) metrics. We can also use well-established physical theory to predict[1] many aspects of the evolution of the Universe. We will examine this model and the evidence supporting it in more detail below.

Like every other successful scientific theory, this model has enabled us to ask more sophisticated questions, many of which we are not yet able to answer. Specifically, much contemporary research in theoretical cosmology centers on two issues:

1. Extrapolating the big bang model to extremely early times, using (sometimes quite speculative) ideas from particle physics.

2. Understanding and modeling the evolution of initially small inhomogeneities, which

---

[1] Or perhaps one should say "retrodict."

are thought to have grown via gravitational instability into the galaxies and large-scale structure we see today.

The two topics are closely related in that one might hope to use ideas from early-Universe cosmology to determine the initial conditions for the growth of perturbations. In particular, the various inflationary theories of the early Universe make definite and testable predictions about the initial spectrum of perturbations.

A breathtaking amount of theoretical work has been performed in these areas, and a substantial number of methods have been found for confronting these theories with observation. In this dissertation we will be concerned with the study of the cosmic microwave background (CMB) radiation, which has proved to be one of the most powerful tools for testing these theories, although we shall have frequent occasion to touch on the many other sorts of observational evidence relevant to cosmology. As we shall see below, the CMB is a relic left over from a very early epoch. The typical microwave background photon is thought to have last scattered off an electron at a time when the Universe was approximately one thousand times smaller than it is today.[2] The CMB is therefore an invaluable source of information about the evolution of the Universe. In particular, fluctuations in the observed temperature of the CMB at different points on the sky contain the imprints of small fluctuations in the distribution of matter at this early epoch. The CMB therefore provides a unique window into the Universe's younger days. It can be used to test ideas about the early Universe; we shall see that inflation in particular makes quite definite predictions about the statistical properties of the observed CMB fluctuations. Furthermore, by comparing the small fluctuations imprinted in the CMB with the much larger density variations we see in the Universe today, we can hope to test the gravitational instability hypothesis, and perhaps to make discoveries about the nature and amount of matter in the Universe.

The remainder of this chapter is organized as follows. In Section 1.2 we give a brief overview of the hot big bang model. Section 1.3 contains a brief summary of the theory

---

[2] The number 1000 depends upon the some assumptions about the ionization history of the Universe. In particular, if the Universe underwent early reionization, then the epoch of last scattering may be somewhat more recent. In any case, however, the microwave background photons are still much older than any objects we can observe directly. We can be quite confident that the CMB originates from further away than clusters of galaxies at fairly high redshift, because we have observed the Sunyaev-Zel'dovich effect, *i.e.*, a spectral distortion to the CMB caused by inverse Compton scattering, in such clusters [17, 97, 192]. Furthermore, the fact that the CMB has a blackbody spectrum suggests that it was produced by an optically thick source. Such a source must be further away than the most distant radio galaxies at redshifts of a few; if it were closer, it would obscure the radio galaxies.

of CMB anisotropy. In the final section, we provide a guide to the remainder of this dissertation.

1.2 The Hot Big Bang Model

We will not attempt to give a comprehensive review of the big bang model here. Such a review is a monumental task, and volumes have been written in attempts to accomplish it. We refer the reader who is interested in reviews of the big bang model in general to the books by Peebles [134, 136, 138], Kolb & Turner [108], and Padmanabhan [132] for such a detailed treatment, and to Longair [119] and Peebles *et al.* [139] for briefer overviews. The reader in search of a review of current topics in cosmological perturbation theory and large-scale structure may wish to look at the reviews by Efstathiou [57, 58], White [190], and Ostriker [131]. In this section we will briefly summarize some important aspects of the theory and some of the observational support for it, concentrating on those aspects that are relevant to the rest of this dissertation.

When Scheuer claimed in 1963 that cosmology was a discipline containing only $2\,{}^1\!/_2$ facts, he was referring to the following:

1. The night sky is dark.

2. The galaxies are receding from one another with speeds that are proportional to their distances.

$2\,{}^1\!/_2$. Some classes of astronomical objects (*e.g.*, radio galaxies) appear to have been different at earlier epochs from the way they are today.

In the intervening three decades, the observational situation has improved considerably: newer and more powerful optical telescopes have been built, and, more importantly, we have learned to observe the sky over a much broader set of wavelengths. Even still, in his rather conservative review [119], Longair extends Scheuer's list to a mere nine facts.

Even Scheuer's meager list permits some interesting conclusions. The observation that the night sky is dark leads to what is generally known as *Olbers's paradox* [130], although it had been around for at least a century before Olbers got to it [78]. If we suppose that the Universe is spatially infinite, has existed for an infinite amount of time, and contains a homogeneous, static distribution of stars (or galaxies), then every line of sight should end on the surface of a star, so that the surface brightness of the night sky should be the same as that of a typical star. The very fact that the sky is dark at night forces us to conclude that at least one of these assumptions is false.

The big bang model has two independent ways of resolving Olbers's paradox. First, the Universe has a finite age. Light has had only a certain amount of time to travel since the beginning, so even if the Universe is infinite, the volume that is visible to any particular observer is finite.[3]

The second resolution of Olbers's paradox arises from the second fact on Scheuer's list, namely Hubble's [92] observation that the light from galaxies is redshifted in proportion to their distance. Distant galaxies have extremely large redshifts, and the light from them is therefore shifted to unobservably low energies. Of course, the discovery of the Hubble redshift law did a great deal more than simply resolve Olbers's paradox. It led to a radical new conception of the Universe, in which the galaxies are all flying apart from one another.

Hubble's discovery of the expanding Universe led to two competing theories. The steady state theory proposed that the Universe, on average, does not change with time. In order to reconcile this postulate with the observed expansion, one must propose that new matter is constantly being created to keep the density constant. The steady state cosmology is inconsistent with Scheuer's fact $2\,{}^1\!/_2$, which shows that in at least some ways the Universe is not invariant under translations in time.[4] The model is also unable to explain the cosmic microwave background radiation and has difficulty explaining the high cosmic helium abundance. As we shall see below, its rival, the big bang model, explains both of these phenomena in an extremely natural way.

In the big bang model, we abandon the "perfect cosmological principle" that the Universe is invariant under time translations. We take at face value the observation that the Universe is expanding and we presume that there is no mysterious process creating matter *ex nihilo*. It then follows that the Universe is getting larger and less dense with the passage

---

[3] Interestingly enough, one of the first people to propose this resolution of Olbers's paradox was Edgar Allen Poe, who suggested that "the distance of the invisible background is so immense that no ray from it has yet been able to reach us at all" [142].

[4] In 1963, Scheuer considered this to be half a fact. It is by now certainly a full-fledged fact. There are several independent lines of evidence suggesting that various populations of objects have evolved over cosmological time scales. For example, the morphology of radio galaxies varies systematically with redshift [37], and high-redshift galaxies tend to be bluer than their nearby cousins [42]. But perhaps the most visually striking evidence that galaxies evolve with time comes from recent Hubble Space Telescope images of very distant galaxies [55]. In these remarkable images the morphological difference between galaxies at high redshift and nearby galaxies are unmistakable.

of time. In particular, at very early times the density must have been extraordinarily high, and in fact one can extrapolate all the way back to a time when the density was infinitely high. This moment is dubbed "the Big Bang," although a naïve extrapolation all the way back to this moment is presumably not warranted, and in fact we know nothing at all about that moment itself (if there even was such a moment).

If we throw in a few more assumptions, we can begin to construct a genuine physical theory of cosmology. First, let us assume that on average the distribution of matter is homogeneous and isotropic. We know, of course, that this is false if we look on fine enough scales: the Earth is many times more dense than the interstellar medium in our own Galaxy, to say nothing of the intergalactic medium. However, the assumption does seem to be correct if we average over a scale of, say, several hundred megaparsecs. Until recently, hard evidence for large-scale homogeneity was somewhat scanty; in particular, surveys of the distribution of galaxies always seemed to reveal clustering on scales as large as the scale of the survey. That has not stopped visionaries,[5] including Einstein [61], from making this assumption, however. Nowadays, the evidence supporting the assumption seems solid [138].

If we combine the assumptions of homogeneity and isotropy with Einstein's general relativity, we can describe the geometry of the Universe in great detail. Friedmann [63] observed that a homogeneous and isotropic solution to the Einstein equations must necessarily have a metric of the form

$$ds^2 = dt^2 - R^2(t)\left(dr^2 + S^2(r)(d\theta^2 + \sin^2\theta\, d\varphi^2)\right), \tag{1.1}$$

where

$$S(r) = \begin{cases} \sin r \\ r \\ \sinh r, \end{cases} \tag{1.2}$$

if the spatial curvature is positive, zero, or negative respectively. The spatial part of this metric describes a three-dimensional surface of constant curvature, either a three-sphere, Euclidean space, or hyperbolic space.[6] $R(t)$ is the time-dependent overall scale of the

---

[5] We call them visionaries because they turned out to be right. Otherwise, we would have a different name for them.

[6] If we choose, we can replace the assumption of isotropy with the weaker assumption of local isotropy. (Specifically, we demand not that the elements of the group of rotations about a point be isometries of the whole spacetime manifold, but only that they act isometrically on a neighborhood of each point.) In this case, space must still be locally isometric to one of these three manifolds, but the global topology may be nontrivial. Such nontrivial topologies exist whether the Universe is open, flat, or closed. In general, we will be concerned only with models that have the standard simply-connected topology (that of the three-sphere

solution. In models with spatial curvature, it is the curvature radius; it is an arbitrarily-normalized scale factor in spatially flat models.

If we know the stress-energy tensor of the matter in the Universe, then the Einstein equation gives a differential equation that can be solved for $R(t)$. To be specific, one frequently assumes that the matter in the Universe is a perfect fluid characterized by some time-dependent energy density ρ and pressure P. With these assumptions, the time evolution of the scale factor is described by the Friedmann equation,

$$\left(\frac{\dot{R}}{R}\right)^2 + \frac{k}{R^2} = \frac{8\pi G\rho}{3}, \tag{1.3}$$

where k takes on the values $+1$, 0, or -1 depending on whether the spatial curvature is positive, zero, or negative. This equation alone is insufficient to specify the dynamics: we need to know how ρ depends on time. Local energy conservation requires that the density and pressure satisfy the relation $d(\rho R^3) = -P\,d(R^3)$. This relation gives the familiar results $\rho \propto R^{-3}$ for pressureless dust, and $\rho \propto R^{-4}$ for radiation (with $P = \rho/3$). For the case of vacuum energy (also known as a cosmological constant), ρ should be independent of R, and so vacuum energy must be described by a negative pressure $P = -\rho$.

We can draw a number of useful conclusions by inspecting the Friedmann equation (1.3). As long as ρ decays faster than R^{-2}, the curvature term k/R^2 in equation (1.3) is negligible in comparison to the matter term at sufficiently early times. Furthermore, the sign of the curvature is simply the sign of $\Omega - 1$, where the density parameter Ω is the ratio of the density ρ to a certain critical density,

$$\Omega \equiv \frac{\rho}{\rho_{\rm crit}} \equiv \frac{\rho}{3H^2/8\pi G}, \tag{1.4}$$

and $H = \dot{R}/R$ is the Hubble parameter.

The case where $k = 0$ (and hence $\Omega = 1$) is of particular interest, both because it is bound to be a good approximation at early times, and because we have some reasons to suspect that we may live in such a flat Universe today. The Friedmann equation has particularly simple solutions in this case. In particular, $R \propto t^{2/3}$ if the Universe is filled with pressureless dust, and $R \propto t^{1/2}$ if the Universe is dominated by relativistic matter so that $P = \rho/3$. Finally, if the Universe is dominated by vacuum energy (or equivalently if it has

---

if the Universe is closed, otherwise that of $\mathbf{R}^3$. Models with nontrivial topology are described briefly in Appendix C.

a large cosmological constant Λ), then the scale factor grows exponentially: $R \propto \exp(Ht)$, with H determined by setting $\Omega = 1$ in equation (1.4).[7]

We now turn from the dynamics of space itself to the behavior of the matter within it. As the Universe expands, its contents cool adiabatically so that the temperature is inversely proportional to the scale factor R. Our belief that the Universe did in fact cool from an initial hot state is based primarily on two sorts of observational evidence. The first class of evidence, about which we will have little to say here, consists of measurements of the relative abundances of isotopes of the light elements hydrogen, helium, and lithium. One can predict the isotopic abundances produced during the hot early Universe (at time $t \sim 1$ minute). The results are in satisfactory agreement with observations, although the process of converting observed abundances to "primordial" ones is far from trivial.[8] Furthermore, the high abundance of ^{4}He (about 23% by mass) is quite hard to explain in any way other than by fusion in the early Universe. For further details on the current status of big bang nucleosynthesis, see [159, 181] and references therein.

The second piece of evidence supporting the hypothesis of a hot early Universe is the existence of the microwave background radiation itself. In the hot early Universe, matter and radiation should have been in thermal equilibrium. The early Universe should therefore have contained photons with a blackbody spectral distribution. Eventually, as the temperature and density decreased, the reactions that coupled the photons to the matter began to occur too infrequently to maintain thermal equilibrium; however, even after these reactions had stopped completely, the photons retained their blackbody distribution, with temperature decreasing as R^{-1}. (Each individual photon is redshifted in wavelength in direct proportion to R, while the photon number density simply decreases in inverse proportion to the volume element R^3. These two effects conspire to maintain the blackbody spectrum.) The big bang model therefore predicts that the Universe should be filled with blackbody radiation today.

This blackbody radiation was inadvertently discovered in 1964 by Penzias and Wilson [141]. It was immediately recognized as a relic from the early Universe by Dicke *et al.* [52],

---

[7] We are restricting our attention here to solutions in which $\dot{R} > 0$, since we know that the Universe is expanding. The Einstein equation is invariant under time reversal, so we could obtain contracting solutions by replacing t by $-t$ in these solutions.

[8] The theoretical predictions depend on one unknown parameter, the baryon-to-photon ratio (or equivalently, since we know the current photon density quite well, the baryon density today). Since we have several different isotopes to use (D, ^{3}He, ^{4}He, and ^{7}Li, all measured relative to H), we can simultaneously fit to this parameter and test the consistency of the whole theory.

who had just embarked on a project to search for it. The spatial distribution and spectrum of this blackbody radiation have been subjects of intense study over the succeeding decades. According to the FIRAS instrument aboard the COBE satellite, the spectrum matches that of an ideal Planckian black body to the level of about one part in 10^5 [123].[9] The blackbody temperature is $T = 2.726 \pm 0.010$ K (95% confidence). The near perfection of the Planckian spectrum argues strongly for a cosmological origin.

The nearly perfect isotropy of the microwave background radiation also suggests that it is a relic of the early Universe. Any local source would be expected to show some anisotropy, since the distribution of matter nearby (both within our Galaxy and in its neighborhood) is quite far from uniform. The microwave background, however, is isotropic to the level of about one part in 10^5 over a wide range of angular scales [188], apart from a dipole anisotropy at the level of one part in 10^3, which is interpreted as the effect of our motion with respect to the CMB center-of-momentum frame [40, 106, 160]. Of course, the isotropy of the CMB is not quite perfect, as was first discovered by Smoot *et al.* in 1992 [163]; the characterization of this slight anisotropy forms the primary concern of this document.

On the basis of this evidence (and more), nearly all cosmologists have accepted the hot big bang model and have turned their attention to a number of open questions within the model. In particular, a great deal of attention has focused on the evolution of inhomogeneities in the Universe. The big bang model proposes that the Universe is nearly homogeneous on large scales, and the nearly perfect isotropy of the CMB suggests that the Universe was very close to homogeneous over a wide range of scales at the time of last scattering. However, we know that the Universe today is quite far from homogeneity on scales less than about 100 Mpc: matter[10] is lumped into galaxies, which are in turn organized into groups and clusters, and even on the largest scales into superclusters and "great walls." The mechanism by which this large-scale structure formed has been a topic of great interest in recent decades.

There is widespread belief that the solution to this puzzle is that structure grew via gravitational instability. If the early Universe contained slight departures from perfect homogeneity, then these density fluctuations could have grown under the influence of gravity

---

[9] Remarkably, the microwave background is at least as good a black body as anything we know how to produce in a laboratory.

[10] Luminous matter, anyway.

to produce the structure we see today. This hypothesis makes a great deal of qualitative sense, and it is therefore of considerable interest to test it quantitatively. The CMB provides a powerful probe of this hypothesis: if the gravitational instability hypothesis is correct, then the initial "seed" fluctuations should leave their imprint as spatial variations in the CMB temperature we observe today.

Although the process of making detailed quantitative predictions of CMB anisotropy is quite demanding, the physical picture is easily grasped. In the standard big bang model, the photons and baryons are quite tightly coupled until a redshift of around 1000. At this time the rate of interaction between photons and electrons drops dramatically, primarily because of the formation of neutral hydrogen atoms. The mean free path of a photon is extremely large thereafter, and indeed most photons propagate freely from that epoch until today. When we look at the microwave background, we are therefore seeing a snapshot of the Universe at the time of last scattering. As we will see below, several distinct physical effects, both on the surface of last scattering and in the intervening space, leave their imprints on this snapshot. (If the Universe underwent early reionization, then a large fraction of photons do scatter again after redshift 1000. As we will see, this has the effect of erasing some of the primordial fluctuations.)

Early work in the field [140, 157, 158] was primarily concerned with a Universe filled with ordinary baryonic matter. In these models the seed fluctuations required to explain structure formation predicted CMB anisotropy at the level of about one part in 10^3. The realization that the actual amplitude of anisotropy was below these levels was one of the motivations for consideration of cosmological models dominated by nonbaryonic dark matter.

In order to test the gravitational instability hypothesis, one needs to make a number of guesses about initial conditions. In particular, one must specify the density and composition of the Universe, as well as the statistical properties of the initial density fluctuations, before it is possible to make quantitative predictions to compare with observations. If there is no fundamental theory available to make definite predictions on these subjects, one sometimes relies on simple educated guesses. For example, one might guess on grounds of "naturalness" or æsthetics that the Universe should be spatially flat,[11] and in the absence of a better

---

[11] The idea that $\Omega = 1$ is in some sense "natural" is usually based on the fact that 1 is an unstable fixed point for Ω. If Ω differs slightly from one at some early time, then $|\Omega - 1|$ grows extremely rapidly. Therefore, since we know that Ω is fairly close to one today, we can conclude that it must have been *extremely* close to one at early times. If one is going to go to the trouble of "fine-tuning" Ω to be $1 \pm \varepsilon$ with $\varepsilon \ll 1$ at early

idea one might guess that the initial density fluctuations were Gaussian distributed with a smooth power-law power spectrum. The remaining free parameters (*e.g.*, the normalization and spectral index of the power spectrum) would then be determined by fitting to the data.

During the 1980's these guesses began to be supplanted with theoretical models that attempted to predict the initial conditions for structure formation from early-Universe physics. In particular, models in which the early Universe went through an inflationary period of rapidly accelerating expansion became quite popular. (See [127] for a review of inflationary cosmology.) These models solve the so-called horizon problem; that is, they explain how widely separated points in the Universe "know" to have the same temperature and density. Inflation also explains why Ω is as close to one as it is: during an inflationary epoch, the quantity $|\Omega - 1|$ decreases extremely rapidly.[12]

In addition to solving the horizon and flatness problems, inflation provides a natural physical mechanism for generating initial density perturbations. Quantum fluctuations during the inflationary epoch are "stretched" by the expansion to become density perturbations today. The fluctuations generated in these models are Gaussian with a power spectrum that is close to the scale-invariant form $P(k) \propto k$. (Note that P no longer denotes pressure.) Inflationary models therefore go a long way towards completely specifying the initial conditions for the structure formation problem. In particular, we can use these models to make definite predictions for the statistical properties of the CMB anisotropy as well as for the present matter distribution.

1.3 Microwave Background Anisotropy as a Cosmological Probe

In this section we will give a brief introduction to the theory of CMB anisotropy. We will not attempt a comprehensive review of this vast subject; the interested reader is referred to review articles by Bond [23] and White *et al.* [188] and to the excellent Ph.D. thesis of Wayne Hu [85]. We will instead simply describe qualitatively the relevant physical effects.

---

times, one might as well go all the way and suppose that it was exactly one. The present author has never been overwhelmed by the persuasiveness of this argument.

[12] A more physically pleasing way to say this is that the curvature radius grows to be much larger than the present horizon size, so no spatial curvature is detectable today. It is, however, possible to contrive inflationary models in which spatial curvature is not negligible [29, 121, 144, 197].

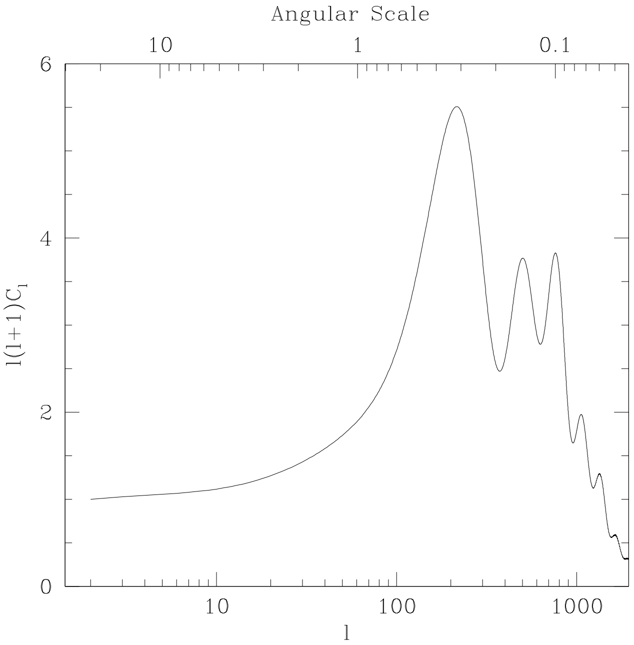

Figure 1.1: Angular power spectrum for a cold dark matter model

The angular power spectrum C_l is shown for a particular inflation-inspired model. The angular scales indicated at the top of the figure are given by the simple relation $\theta = 60^\circ/l$. The overall normalization of the power spectrum is arbitrary.

The foundations of the theory of CMB anisotropy date back almost to the discovery of the background radiation itself. Pioneers in the field include Sachs & Wolfe [149], Silk [157, 158], Sunyaev & Zel'dovich [172], and Peebles & Yu [140]. Over the intervening decades we have learned to make extremely precise predictions of the statistical properties of the CMB anisotropy in particular models. Several groups, beginning with Bond & Efstathiou [24, 25], have written programs that numerically integrate the Boltzmann equation for the coupled matter-radiation fluid to obtain the CMB angular power spectrum. These programs are now accurate to levels of a few percent [91]. There are also surprisingly accurate analytic approximations that provide valuable insight into the physical mechanisms producing the anisotropy [85, 88].

Figure 1.1 shows the angular power spectrum C_l for a typical inflation-inspired model. The particular model shown has spectral index $n = 1$, Hubble parameter $H_0 =$

50 km/s/Mpc, and baryon fraction $\Omega_B = 0.06$. The remaining 94% of the matter is nonbaryonic cold dark matter. We will discuss angular power spectra in more detail in Chapter 2; for the moment all we need to know is that C_l is the variance of the coefficient a_{lm} of a spherical harmonic expansion of the temperature fluctuation ΔT on the sky. As indicated in the figure, large values of l correspond to fluctuations on small angular scales and *vice versa.* Our primary concern in this dissertation will be with the COBE DMR data. Since the DMR has a resolution of 7° (FWHM), it is insensitive to fluctuations on small angular scales, and so our primary concern will be with values of l less than about 30.

Figure 1.1 contains a surprising amount of structure, especially considering that the primordial inflation-generated density fluctuations have a featureless power-law power spectrum. As we will see in the remainder of this section, the reason for this is that several distinct physical effects conspire to produce the total anisotropy power spectrum.

For the large scales probed by COBE, the most important mechanism for producing anisotropy is the Sachs-Wolfe effect [149]. This is simply the gravitational redshift or blueshift suffered by the CMB photons between the surface of last scattering and here. In the simplest case, where the gravitational potential does not evolve in time, the fractional energy shift of each photon is simply the difference in gravitational potential between emission and reception. If the gravitational potential does change with time, then there is an additional energy shift obtained by integrating the derivative of the potential along the photon's world line. This additional shift often goes by the name of the integrated Sachs-Wolfe effect, and in the case where the changing potential is due to the formation of nonlinear structures it may be called the Rees-Sciama effect. Whatever name one gives to the effect, credit for its discovery belongs to Sachs and Wolfe.

If we assume that $\Omega = 1$, that the Universe is dominated by nonrelativistic matter, and that all departures from homogeneity are small enough for linear perturbation theory to be a good approximation, then the gravitational potential does not evolve in time and there is no integrated Sachs-Wolfe effect. As we will see in Chapter 2, the Sachs-Wolfe effect in this case gives a power spectrum $C_l \propto 1/l(l+1)$ if the primordial density power spectrum is of the Harrison-Zel'dovich form $P(k) \propto k$. This is the explanation of the flat portion of the power spectrum at large angular scales in Figure 1.1. However, the assumption of matter domination is not quite correct at the time of last scattering: the energy density of photons and relativistic neutrinos is not entirely negligible. There is therefore a small integrated Sachs-Wolfe contribution, which is the primary cause of the slow rise in the power spectrum at low l.

The prominent peaks in the power spectrum on angular scales $\sim 0\overset{\circ}{.}5$ are generally referred to as "Doppler peaks," although the term is something of a misnomer [88]. They are caused by acoustic waves propagating through the photon-baryon fluid during the time leading up to last scattering. Only waves with wavelengths smaller than the horizon distance at the time of last scattering have time to develop; this imposes the characteristic angular scale of about half a degree. The sequence of peaks is akin to a harmonic series: the first peak corresponds to waves that are just reaching their first antinode at the time of last scattering, while subsequent peaks are produced by higher-frequency modes that have oscillated more than once.

In explaining both the Sachs-Wolfe plateau and the "Doppler" peaks, we have implicitly had in mind a fairly simplistic model: we have imagined that the photon-baryon fluid remains tightly coupled right up until the moment of last scattering, at which moment the photons are instantaneously and simultaneously released. This model is good enough for explaining relatively large-scale features, but if we want to understand the precipitous decline in the angular power spectrum around $l \sim 1000$, we need to do slightly better. Around the time of last scattering, the photon mean free path increases dramatically.[13] Perturbations in the photon density are therefore erased by diffusion: photons wander out of overdense regions and into underdense regions. This process, which is known as "collisional damping" or "Silk damping," erases fluctuations on scales less than a few megaparsecs [158] and explains the decrease in power at small scales in Figure 1.1.

The same qualitative features — the Sachs-Wolfe plateau, the "Doppler" peaks, and Silk damping — occur in a wide range of models. Indeed, by measuring the locations and amplitudes of these various features, we can hope to determine cosmological parameters such as the baryon fraction and the Hubble constant. See [85] for details about the manner in which these features depend on cosmological parameters. In all models in which recombination occurs on schedule at $z \simeq 1000$, the effects of Silk damping are completely unobservable on the large angular scales probed by COBE; in fact, even the acoustic peaks are largely irrelevant on COBE scales in these models. Our primary concern will therefore be with the Sachs-Wolfe effect (both integrated and otherwise), although we will comment briefly in Chapter 5 on reionized models, in which other effects are quite important.

---

[13] Indeed, after last scattering, the mean free path is greater than the horizon distance; after all, that's what "last scattering" means.

The effects we have described in this section by no means form an exhaustive list of sources of CMB anisotropy, although they are the most important contributors on large angular scales. We have not considered models in which the Universe contains topological defects such as strings or texture; these defects impose their own characteristic forms of CMB anisotropy. In addition, long-wavelength gravitational waves can produce anisotropy *via* the Sachs-Wolfe effect; we will have more to say on this subject in Chapter 5. Finally, everything we have described is based on linear perturbation theory; on small angular scales, and especially in reionized models, higher-order effects can be important [87].

We have not yet mentioned the vexing issue of foreground contamination. It can be quite difficult to be certain that the anisotropy we see in the sky is due to the CMB and not some contaminant within our own Galaxy. In particular, thermal emission from dust and synchrotron and free-free emission can all be significant sources of confusion for CMB observations. Fortunately, these sources have significantly different spectral signatures from CMB anisotropy, and so it is in principle possible to separate them out. The COBE DMR experiment is thought to be largely free from Galactic contamination for Galactic latitudes greater than 20°. This hypothesis is supported by the spectral characteristics of the signal, as well as the fact that the overall amplitude of fluctuations does not appear to vary with latitude over this portion of the sky [12, 13, 163]. The fact that the DMR data appear to be consistent with experiments at widely separated frequencies [66, 183] also strongly suggests that the DMR is seeing real CMB anisotropy rather than foreground contamination. The status of foreground contamination in experiments on smaller angular scales is far less clear [28, 145].

1.4 A Guide to This Dissertation

The remainder of this dissertation is organized as follows. Chapter 2 contains a technical introduction to the subject of CMB data analysis. We begin by describing the assumptions we will make about the statistical properties of the CMB anisotropy. We derive the Sachs-Wolfe formula for the CMB anisotropy power spectrum due to gravitational potential fluctuations and comment on the window functions that filter the anisotropy seen by various experiments. The remainder of the chapter is concerned with the COBE DMR experiment. We briefly describe the instrument and the data set, and establish some convenient notation; we conclude with a discussion of the controversial question of what to do with the seemingly anomalous COBE quadrupole.

Chapter 3 describes some simple "toy models" for CMB anisotropy analysis. Although the models described are designed to be similar in some ways to the DMR data, our purpose in this chapter is not to derive precise quantitative results that can be applied to any particular experiment. Rather, we aim simply to assess what sort of questions can fruitfully be asked of a data set like that of the DMR. In particular, we observe that in the absence of complete sky coverage it is impossible to estimate individual spherical harmonic coefficients. Fortunately, the cosmologist's primary concern is estimating the angular power spectrum, a task which need not be seriously compromised by incomplete sky coverage.

In Chapter 4 we describe in detail a method for extracting likelihood information from experiments such as the DMR. The Karhunen-Loève transform described therein is essentially a method for "compressing" the data by expanding it in a set of basis functions and throwing away those modes that are dominated by noise. We perform the transform on the two-year DMR data and give some preliminary results, and we also describe some tests of the method.

Chapter 5 contains the results of confronting the Karhunen-Loève transformed DMR data with a number of cosmological models, with particular emphasis on cold dark matter models both with and without a cosmological constant. We provide maximum-likelihood power-spectrum normalizations as well as relative likelihoods for these models. We also place Bayesian constraints on the value of the cosmological constant Λ in simple models.

In the final chapter we derive the Wiener filter and apply it to the DMR data. Our goal in this chapter is to "clean up" the noise in the data and identify statistically robust features in the maps. In addition, we use the filtered maps to make predictions for the Tenerife experiment, which probes slightly smaller angular scales than COBE.

The reader who does not wish to read the entire document may be interested to know that Chapters 2 and 3 may each be read independently. Chapters 4, 5, and 6 rely on some material in Chapter 2, especially Section 2.6, and Chapter 5 depends fairly heavily on material in Chapter 4.

Chapter 2

An Introduction to Microwave Background Data Analysis

> Our experience with microwaves has followed a familiar contemporary pattern: inadequate response to great expectations.
>
> Irma S. Rombauer and Marion R. Becker
> *The Joy of Cooking*

2.1 Introduction

Although Rombauer & Becker's interests are culinary rather than cosmological [147], their view may have been shared by many cosmologists several years ago. The enormous potential of microwave background experiments to probe the evolution of the Universe had been well known for years, but all attempts to discover deviations from a perfectly uniform blackbody background had failed.

With the discovery of anisotropy by the COBE satellite [163] and the subsequent flurry of reported detections on smaller angular scales (see [188] for a review), the situation changed dramatically. The CMB now provides the best way of normalizing the spectrum of fluctuations in models of structure formation [31, 185], and as the data on degree scales improve in accuracy, the shape of the CMB power spectrum will provide powerful constraints on models. Over the last few years, therefore, much attention has been focused on developing techniques for extracting cosmological information from CMB anisotropy experiments. Such techniques are the primary subject of this dissertation.

In this chapter we will establish the statistical framework for the problems in CMB data analysis that will be tackled in subsequent chapters. In Section 2.2 we will discuss the predicted statistical properties of CMB anisotropy, including a description of Gaussian random fields on the sphere. In Section 2.3 we comment briefly on the methods by which the angular power spectrum corresponding to a particular theory may be computed, and derive the important Sachs-Wolfe angular power spectrum formula (2.29). Section 2.4 describes the window functions through which the anisotropy ΔT is filtered by particular experiments. In Section 2.5 we describe the properties of the COBE DMR data, whose analysis will comprise most of the rest of this dissertation. Section 2.6 establishes some useful notation, and Section 2.7 addresses an important technical issue having to do with the possibility that the COBE quadrupole may be contaminated.

2.2 The Statistics of Microwave Background Anisotropy

Our primary concern in this dissertation will be with theories in which the CMB anisotropy is a realization of a Gaussian random field. This is likely to be a fairly good assumption. Theories based on inflation predict quite unambiguously that the CMB anisotropy should be described by Gaussian statistics [112, 188]. Some theories, such as those based on cosmological defects, predict non-Gaussian CMB anisotropy [27, 98, 179, 180]; however, even these theories predict only very small departures from Gaussian statistics on large angular scales [14, 26, 41, 56]. (This is essentially due to the central limit theorem: the large-angle anisotropy in these models is a superposition of many independent fluctuations, and so the distribution tends towards a normal distribution.) Since we will be concerned primarily with the COBE DMR data, which probe only scales greater than about 7°, we do not expect to see any effects of non-Gaussian structure even if one of these defect models proves to be correct. In any case, the DMR data have been processed in various ways to test for any sign of departure from Gaussian statistics with null results [82, 107, 163].

When we say that the CMB anisotropy ΔT is a realization of a Gaussian random field, we mean that for any sequence of points $(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \ldots, \hat{\mathbf{r}}_n)$ on the sky, the n-dimensional vector

$$(\Delta T(\hat{\mathbf{r}}_1), \Delta T(\hat{\mathbf{r}}_2), \ldots, \Delta T(\hat{\mathbf{r}}_n)) \tag{2.1}$$

is a random variable drawn from a multivariate Gaussian distribution [5]. If we specify the mean $\langle \Delta T(\hat{\mathbf{r}}) \rangle$ of this distribution at all points $\hat{\mathbf{r}}$, and the covariance matrix $\langle \Delta T(\hat{\mathbf{r}}_1) \Delta T(\hat{\mathbf{r}}_2) \rangle$

at all pairs of points, then we know everything there is to know about the statistical properties of the anisotropy.[1]

The assumption of Gaussian statistics is quite restrictive. Without this assumption, we would have to treat the entire probability distribution as an unknown function, to be predicted from a theory or estimated from the data. With the assumption of Gaussianity, only the means and covariances are unknown, leaving us a much smaller parameter space to explore. Another way to say this is to note that a Gaussian distribution is determined entirely by its first two moments, while in general one needs to specify the third moment $\langle \Delta T(\hat{\mathbf{r}}_1)\Delta T(\hat{\mathbf{r}}_2)\Delta T(\hat{\mathbf{r}}_3)\rangle$ and indeed the entire sequence of higher moments to identify a distribution.

In most cases, we can assume that the CMB anisotropy is homogeneous and isotropic as well. This is true in cosmological models in which there is no preferred direction on the sky. All "mainstream" cosmologies satisfy this property, although some multiply connected (topologically "small") models and models with large-scale rotation or shear (Bianchi models) do not. (See Appendix C for some remarks about such nonstandard models.) These assumptions mean that the probability distribution of ΔT is invariant under the group of isometries of the sphere. Specifically, the mean $\langle \Delta T(\hat{\mathbf{r}})\rangle$ is a constant over the sphere. Furthermore, since ΔT is a fluctuation about the average temperature, this constant must be zero. In addition, the covariance between ΔT at two points must depend only on the angle between those points:

$$\langle \Delta T(\hat{\mathbf{r}}_1)\Delta T(\hat{\mathbf{r}}_2)\rangle = C(\theta), \tag{2.2}$$

where

$$\theta = \cos^{-1}(\hat{\mathbf{r}}_1\cdot\hat{\mathbf{r}}_2) \tag{2.3}$$

is the angle between $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$. With these assumptions, the statistical properties of the anisotropy are described entirely by the correlation function $C(\theta)$. It is often convenient to expand C in Legendre polynomials:

$$C(\theta) = \frac{1}{4\pi}\sum_{l=0}^{\infty}(2l+1)\,C_l\,P_l(\cos\theta). \tag{2.4}$$

In ordinary Euclidean space, the simplest way to characterize a homogeneous, isotropic Gaussian random field is generally in terms of its Fourier transform. Distinct Fourier

---

[1] Here, and throughout this dissertation, the angle brackets $\langle\cdots\rangle$ denote an average over a hypothetical ensemble of Universes. In general, random fields on the sphere are *not* ergodic, and so this ensemble average is not the same as an average of one realization over the sphere.

modes are uncorrelated, and so all of the statistical information is contained in the variances of the individual modes. Because of isotropy, these variances depend only on the amplitude of the wave vector, and so the power spectrum $P(|\mathbf{k}|) \equiv \langle|\tilde{f}(\mathbf{k})|^2\rangle$ is a complete description of the random field f. One can make corresponding statements about random fields on the sphere, which will prove quite useful to us.

We begin by expanding the random field ΔT in spherical harmonics:

$$\Delta T(\hat{\mathbf{r}}) = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\hat{\mathbf{r}}). \tag{2.5}$$

We have omitted the monopole ($l = 0$) term from the sum because ΔT must average to zero over the whole sphere, and we have omitted the dipole ($l = 1$) because it is impossible to distinguish from the much larger kinematic dipole caused by our motion relative to the CMB rest frame[2]. The coefficients a_{lm} are

$$a_{lm} = \int \Delta T(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}})\, d\Omega, \tag{2.6}$$

where $d\Omega = d\cos\theta\, d\varphi$ is the usual volume element on the sphere, and (θ, φ) are the spherical coordinates of $\hat{\mathbf{r}}$. The integral ranges over the entire sphere.

The astute reader might have expected a complex conjugate sign in the above equation. We must explain its absence. Throughout this dissertation, the symbol Y_{lm} will denote a real-valued spherical harmonic. The real-valued spherical harmonics are simply the conventional spherical harmonics with complex exponentials replaced by ordinary trigonometric functions:

$$Y_{lm}(\theta,\varphi) = \begin{cases} Y_{lm}^{(conv)}(\theta,\varphi) & \text{for } m = 0 \\ Y_{lm}^{(conv)}(\theta,0)\sqrt{2}\cos m\varphi & \text{for } 1 \le m \le l \\ Y_{lm}^{(conv)}(\theta,0)\sqrt{2}\sin m\varphi & \text{for } -l \le m \le -1 \end{cases} . \tag{2.7}$$

Here $Y_{lm}^{(conv)}$ is a conventional spherical harmonic as defined by, *e.g.*, Jackson [93] or Abramowitz & Stegun [3]. The real-valued spherical harmonics Y_{lm} satisfy the usual orthonormality condition

$$\int Y_{lm}(\theta,\varphi) Y_{l'm'}(\theta,\varphi)\, d\Omega = \delta_{ll'}\delta_{mm'}. \tag{2.8}$$

The addition theorem for the real-valued spherical harmonics is

$$P_l(\cos\gamma) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_{lm}(\theta,\varphi) Y_{lm}(\theta',\varphi'), \tag{2.9}$$

---

[2] In principle, of course, the CMB anisotropy presumably has an intrinsic dipole component; we simply define ΔT to be the dipole-subtracted anisotropy.

where γ is the angle between the points (θ, φ) and (θ', φ'). This differs from the addition theorem for the conventional spherical harmonics (equation (3.62) of [93]) only through the absence of a complex conjugation.

Since the coefficients a_{lm} are linear in ΔT, they are Gaussian-distributed. Furthermore, since $\langle \Delta T \rangle = 0$, it follows from equation (2.6) that $\langle a_{lm} \rangle = 0$. Thus for a complete description of the probability distribution of the a_{lm}'s, all we need is their covariances. Using equation (2.6), we find that

$$\langle a_{lm} a_{l'm'} \rangle = \int d\Omega\, d\Omega'\, Y_{lm}(\theta, \varphi)\, Y_{l'm'}(\theta', \varphi')\, \langle \Delta T(\theta, \varphi) \Delta T(\theta', \varphi') \rangle \tag{2.10}$$

$$= \int d\Omega\, d\Omega'\, C(\gamma) Y_{lm}(\theta, \varphi) Y_{l'm'}(\theta', \varphi'). \tag{2.11}$$

Using equation (2.4) and the addition theorem (2.9), this becomes

$$\langle a_{lm} a_{l'm'} \rangle = \sum_{l'',m''} C_{l''} \int d\Omega Y_{lm}(\theta, \varphi) Y_{l''m''}(\theta, \varphi) \int d\Omega' Y_{l'm'}(\theta', \varphi') Y_{l''m''}(\theta', \varphi'). \tag{2.12}$$

So, using the orthonormality relation (2.8),

$$\langle a_{lm} a_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'}. \tag{2.13}$$

Note that since $a_{lm} = 0$ for $l \leq 1$, $C_0 = C_1 = 0$. To summarize, the a_{lm}'s are independent Gaussian random variables with zero mean and variances C_l. We can completely specify the probability distribution of ΔT by simply giving the sequence C_l. The C_l's are generally known as the angular power spectrum.

At this point it may be wise to comment on the subject of "cosmic variance." Cosmologists work with data sets that are restricted to a sample size of one. We can observe only one Universe, and we observe it from only one vantage point. This unfortunate fact restricts our ability to extract information about the CMB. In particular, there is an unavoidable bound on the accuracy with which we can estimate the angular power spectrum C_l.

Suppose that we had a perfect CMB experiment, which measured ΔT over the entire sky with negligible noise. Then we could determine the spherical harmonic coefficients a_{lm} with perfect accuracy from equation (2.6). The maximum-likelihood estimator of C_l would simply be

$$\hat{C}_l = \frac{1}{2l+1} \sum_{m=-l}^{l} a_{lm}^2. \tag{2.14}$$

Since the a_{lm}'s are independent, identically-distributed Gaussian random variables (for fixed l), $\hat{C}_l$ has a chi-squared distribution with $2l+1$ degrees of freedom. The variance of $\hat{C}_l$ is therefore

$$\frac{\mathrm{Var}(\hat{C}_l)}{\hat{C}_l} = \frac{2}{2l+1}. \tag{2.15}$$

For low values of l this is quite substantial, implying that we will never be able to use the CMB to distinguish among theories that differ only slightly in just their large-scale power spectra.

The term "cosmic variance" refers to the unavoidable minimum uncertainty in equation (2.15). In practice, no experiment ever attains complete sky coverage, and so this limiting fractional uncertainty of $\sqrt{2/(2l+1)}$ in an estimate of C_l is never attained. Cosmologists generally reserve the term "sample variance" for the larger uncertainty that arises due to incomplete sky coverage. For an experiment with complete sky coverage, the sample variance is simply the cosmic variance; if the sky coverage is incomplete, the sample variance grows roughly in inverse proportion to the fraction of sky covered, as one would expect [154, 187].[3] In any case, likelihood-based techniques of the sort that we will describe in Chapter 4 automatically include the effects of sample variance.

2.3 Computing the Power Spectrum

We have seen in Chapter 1 that many different physical effects contribute to CMB anisotropy. Computing the angular power spectrum C_l corresponding to a particular theoretical model is therefore a daunting task. One must follow the evolution of perturbations in the coupled photon-baryon fluid through recombination. In "standard" models with Gaussian initial conditions (such as inflation-inspired models), all of the perturbations are small and one can linearize the equations. Then individual Fourier modes[4] evolve independently, and it is possible to solve the Boltzmann equation numerically to determine the CMB power spectrum. For models with non-Gaussian seeds such as cosmic strings or texture, this procedure does not work. In these cases one must resort to detailed cosmological simulations to make predictions for the CMB. For this reason, the state of the art in CMB prediction is much more precise for inflation-based models than for those based on topological defects.

---

[3] In practice, of course, experiments are noisy, and so the total uncertainty is larger than either the sample variance or the cosmic variance.

[4] Or, if space is not flat, harmonic modes.

We will not embark on a detailed discussion of the computation of power spectra here. The reader with an interest in understanding and solving the Boltzmann equation for the CMB anisotropy is referred to a thorough review by Bond [23]. Other useful sources include [24, 25, 57, 85, 88, 90, 91, 188] and references therein. The reader interested in topological defects is referred to [26] and references therein. We will, however, give a derivation of the contribution of the Sachs-Wolfe effect to the power spectrum of flat adiabatic models. This is the largest contributor to the anisotropy on large angular scales, and so it will be our primary interest when we analyze the COBE DMR data (although, as we shall see, other effects are not entirely negligible even on these scales).

We suppose that $\Omega = 1$ and that the initial density perturbation δ is a homogeneous isotropic Gaussian random field with a power spectrum that is a power law in wavenumber k. In other words, the Fourier transform[5] $\tilde{\delta}_{\mathbf{k}}$ is Gaussian distributed with mean zero and covariance matrix

$$\langle \tilde{\delta}_{\mathbf{k}} \tilde{\delta}^*_{\mathbf{k}'} \rangle = P(k)\delta_D(\mathbf{k}-\mathbf{k}'), \tag{2.16}$$

where

$$P(k) = Ak^n \tag{2.17}$$

and δ_D is the three-dimensional Dirac delta distribution. The gravitational potential is related to the density through Poisson's equation,

$$\nabla^2\phi = 4\pi G\rho_B\delta = \frac{3}{2}H^2\delta, \tag{2.18}$$

where $\rho_B = 3H^2/8\pi G$ is the background density (which is also the critical density), and we are working in the Newtonian gauge.[6] Working in Fourier space, we have

$$\tilde{\phi}_{\mathbf{k}} = -\frac{3}{2}H^2a^2\frac{\tilde{\delta}_{\mathbf{k}}}{k^2}, \tag{2.19}$$

where a is the scale factor. Note that in a flat matter-dominated model, $H \propto a^{-3/2}$ and $\tilde{\delta} \propto a$, so $\tilde{\phi}$ is constant in time. Let us normalize the scale factor to be $a = 1$ at the present, so that

$$\tilde{\phi}_{\mathbf{k}} = -\frac{3}{2}H_0^2\frac{\tilde{\delta}_{\mathbf{k}}}{k^2}. \tag{2.20}$$

---

[5] We denote Fourier transforms with a tilde. Cosmologists generally denote a function and its Fourier transform by the same symbol, relying on context (generally meaning the absence or presence of the letter k) to dispel ambiguity. We find this practice odious.

[6] For an exceptionally lucid discussion of gauges in cosmology, see [57].

The power spectrum associated with the gravitational potential is then

$$P_\phi(k) = \frac{9}{4}H_0^4\frac{P(k)}{k^4} = \frac{9}{4}H_0^4 A k^{n-4}. \tag{2.21}$$

If we consider only the Sachs-Wolfe effect due to gravitational potential fluctuations on the surface of last scattering, the fractional temperature anisotropy is $(\Delta T/T)(\hat{\mathbf{r}}) = \phi(R\hat{\mathbf{r}})/3$ where $R = 2c/H_0$ is the distance to the last scattering surface. Using equation (2.6) and taking a Fourier transform, we have

$$a_{lm} = \frac{1}{3(2\pi)^3}\int d\Omega \int d^3k\, Y_{lm}(\hat{\mathbf{r}}) e^{-iR\mathbf{k}\cdot\hat{\mathbf{r}}}\tilde{\phi}_{\mathbf{k}}. \tag{2.22}$$

Squaring and taking an ensemble average, we find that

$$C_l = \frac{1}{9(2\pi)^6}\int d^3k\, d\Omega_1\, d\Omega_2\, Y_{lm}(\Omega_1) Y_{lm}(\Omega_2)\, P_\phi(k)\, e^{iR\mathbf{k}\cdot(\hat{\mathbf{r}}_2-\hat{\mathbf{r}}_1)}. \tag{2.23}$$

Recall that an exponential can be expanded in spherical harmonics:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_{l,m} i^l j_l(kr) Y_{lm}(\hat{\mathbf{k}}) Y_{lm}(\hat{\mathbf{r}}), \tag{2.24}$$

where j_l is a spherical Bessel function [93]. The two angular integrals in equation (2.23) are simply the spherical harmonic orthonormality relation (2.8), so

$$C_l = \frac{1}{36\pi^4}\int d^3k P_\phi(k) j_l^2(kR) = \frac{1}{9\pi^3}\int dk\, k^2 P_\phi(k) j_l^2(kR). \tag{2.25}$$

Since $P_\phi \propto k^{n-4}$, we have

$$C_l = \frac{1}{\pi^2}H_0^4\int_0^\infty dk\, k^{n-2} j_l^2(kR). \tag{2.26}$$

Performing this integral, we find that

$$C_l = \frac{1}{64\pi^{3/2}}(2H_0)^{n+3}\frac{\Gamma\left((3-n)/2\right)\Gamma\left((2l+n-1)/2\right)}{\Gamma\left((4-n)/2\right)\Gamma\left((2l+5-n)/2\right)}. \tag{2.27}$$

It is customary to normalize power spectra with respect to C_2, since this is the lowest nonzero element. Specifically, one generally defines a quantity

$$Q_{\rm ps} = \sqrt{\frac{5C_2}{4\pi}}, \tag{2.28}$$

which is used to quote the normalization of the angular power spectrum.[7] With this normalization convention, the Sachs-Wolfe angular power spectrum is

$$C_l = \frac{4\pi}{5} Q_{\rm ps}^2 \frac{\Gamma\left((2l+n-1)/2\right)\Gamma\left((9-n)/2\right)}{\Gamma\left((2l+5-n)/2\right)\Gamma\left((3+n)/2\right)}. \tag{2.29}$$

2.4 Window Functions

The bulk of this dissertation will be concerned with estimating the power spectrum C_l associated with the CMB anisotropy ΔT. However, no experiment measures ΔT directly. In every case, the instrument response is the convolution of ΔT with some "window function" depending on the details of the experiment. In this section we will describe how to compute these window functions.

Every experiment has some finite beam-width, and so the anisotropy measured by the experiment is always the convolution of ΔT with some point-spread function. For many experiments, the beam pattern is approximated well by a Gaussian.[8] In this case the instrument response is

$$R(\hat{\mathbf{r}}) = \int d\Omega' \, \Delta T(\hat{\mathbf{r}}') W(\hat{\mathbf{r}}, \hat{\mathbf{r}}'), \tag{2.30}$$

where the window function W is

$$W(\hat{\mathbf{r}}, \hat{\mathbf{r}}') = e^{-|\hat{\mathbf{r}}-\hat{\mathbf{r}}'|^2/2\sigma^2}. \tag{2.31}$$

One frequently characterizes the beam-width in terms of the full width at half-maximum (FWHM). To get the Gaussian beam-width σ one multiplies the FWHM by $(8\ln 2)^{-1/2} = 0.4247$.

The instrument response R can be expanded in spherical harmonics just like ΔT:

$$R(\hat{\mathbf{r}}) = \sum_{l,m} \bar{a}_{lm} Y_{lm}(\hat{\mathbf{r}}). \tag{2.32}$$

---

[7] The quantity $Q_{\rm ps}$ has a number of different names in the literature. In particular, the COBE group calls it Q_{rms-PS}, and the present author used to call it $\langle Q\rangle$ [31, 185]. Both the angle brackets in $\langle Q\rangle$ and the PS (standing for "power spectrum") in Q_{rms-PS} and $Q_{\rm ps}$ are meant to remind the reader that $Q_{\rm ps}$ is an ensemble-average quantity, and need not coincide with the local quadrupole Q defined in equation (2.68) below. We have abandoned our earlier choice of $\langle Q\rangle$ because $Q_{\rm ps}$ is *not* the ensemble average value of Q, but rather $\langle Q^2\rangle^{1/2}$. Incidentally, since Q is χ_5^2 distributed, it is straightforward to check that $Q_{\rm ps}$ differs from the ensemble average of Q by a factor of $(8/3)\sqrt{2/5\pi} = 0.952$.

[8] And for some experiments only one parameter, such as the full width at half-maximum, is known. In such cases, one frequently models the beam as a Gaussian for lack of any better information.

When the window function depends only on the distance $|\hat{\mathbf{r}} - \hat{\mathbf{r}}'|$ as above, the coefficients $\bar{a}_{lm}$ are related in a simple way to the coefficients a_{lm} in the expansion of ΔT. Let us expand W in Legendre polynomials:

$$W(\hat{\mathbf{r}}, \hat{\mathbf{r}}') = \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l+1) W_l P_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}'). \tag{2.33}$$

Then equation (2.30), combined with the addition theorem (2.9), gives

$$R(\hat{\mathbf{r}}) = \sum_{l,m} W_l \int d\Omega' \, \Delta T(\hat{\mathbf{r}}') Y_{lm}(\hat{\mathbf{r}}') Y_{lm}(\hat{\mathbf{r}}) = \sum_{l,m} W_l a_{lm} Y_{lm}(\hat{\mathbf{r}}). \tag{2.34}$$

It follows that

$$\bar{a}_{lm} = W_l a_{lm}. \tag{2.35}$$

The beam-filtered coefficients are obtained by multiplying the original coefficients by the Legendre-transformed beam-pattern.[9] In practice, W_l decreases rapidly for $l \gtrsim 1/\sigma$, signifying that the experiment has little sensitivity to those modes that vary on angular scales smaller than the size of the beam.

The central lobe of the COBE DMR beam pattern is approximated well by a Gaussian with FWHM 7° ($\sigma = 3^\circ\!.0$), but the outer parts of the beam drop off much more slowly than a Gaussian. Fortunately, the beam pattern has been measured and tabulated, as has its expansion in Legendre polynomials [195].

For the DMR sky maps, this is all we need to consider, but for other experiments, particularly those on smaller angular scales, we must also take into account the beam-switching pattern. Many experiments "chop" between two points on the sky and record the difference between these points. The window function for these experiments has a positive and a negative lobe. Some experiments chop between three points, recording the difference between the central point and the average of the other two, in effect measuring the second derivative of ΔT. Even for these beam-switching experiments, it is still often the case that $\bar{a}_{lm}$ is simply a constant multiple of a_{lm}, at least if the axis of the spherical coordinate system is chosen in a suitable way. We will give two examples here; more detailed treatments are available elsewhere [23, 188, 189].

We first consider the MAX experiment [38, 51, 76, 124]. This is a balloon-borne telescope with a beam size $\sigma = 0^\circ\!.2$. The telescope is sinusoidally chopped in azimuth, so

---

[9] The insightful reader has doubtless already realized that this is the spherical equivalent of the convolution theorem for Fourier transforms.

that the time-dependent instrument response is

$$R_t(t) = (\Delta T \star B)(\theta, \varphi + \alpha \sin\tau), \tag{2.36}$$

where the chop amplitude $\alpha \sin\theta = 0\overset{\circ}{.}65$, and (θ, φ) marks the central beam location. $\tau = \omega t$, where ω is the frequency of the chop. B denotes the beam pattern, which can be approximated by a Gaussian of width σ, and the star denotes a convolution. We have chosen a spherical coordinate system in which the zenith is at $\theta = 0$.

The time-series data are demodulated at the chop frequency ω, so that the recorded signal R is the Fourier transform of $R_t(t)$ at frequency ω:

$$R(\theta, \varphi) = N \int_0^{2\pi} d\tau \, (\Delta T \star B)(\theta, \varphi + \alpha \sin\tau) \sin\tau, \tag{2.37}$$

where N is a normalization factor.[10] Expanding ΔT in spherical harmonics, we find

$$R(\theta, \varphi) = N \sum_{l,m} B_l a'_{lm} Y'_{lm}(\theta, \varphi) \int_0^{2\pi} d\tau \, e^{im\alpha \sin\tau} \sin\tau, \tag{2.38}$$

where B_l is a coefficient of the Legendre polynomial expansion of the beam. In the Gaussian approximation,

$$B_l = e^{-\frac{1}{2}\sigma^2 l(l+1)}. \tag{2.39}$$

The primes on a_{lm} and Y_{lm} signify that we have returned for the moment from real-valued to conventional spherical harmonics.[11] This integral is a Bessel function:

$$R(\theta, \varphi) = 2\pi N \sum_{l,m} B_l a'_{lm} Y'_{lm}(\theta, \varphi) J_1(m\alpha). \tag{2.40}$$

Thus, the spherical harmonic coefficients $\bar{a}'_{lm}$ are simply the original coefficients a'_{lm} multiplied by a window function

$$W'_{lm} = 2\pi N J_1(m\alpha) e^{-\frac{1}{2}\sigma^2 l(l+1)}. \tag{2.41}$$

The window function W_{lm} corresponding to the real-valued spherical harmonics is a linear combination of W'_{lm} and W'_{l-m}. The reader is invited to work out the coefficients for himself or herself.

---

[10] The MAX experiment calibrates its receiver by chopping between two loads at uniform temperatures. They therefore choose N so that the response $R(\theta, \varphi)$ is unity when the input signal is a step function $\Delta T(\theta, \varphi') = \vartheta(\varphi' - \varphi)$. Further details may be found in [166].

[11] This is simply to avoid having to treat the three cases $m > 0$, $m = 0$, $m < 0$ individually.

Note that $J_1(x) \propto x$ for small x. This means that the experiment's sensitivity to modes with low $|m|$ is linearly suppressed. This is entirely natural: the experiment is designed to measure the gradient of ΔT in the φ direction, and modes with low $|m|$ vary slowly in this direction.

The other window function we will consider here is a straightforward triple-beam experiment. The Tenerife experiment [45, 79, 182, 183], which we will consider in Chapter 6, is an example of such an experiment. In a triple-beam experiment, the instrument response is the difference between the temperature at a particular point and the average of the temperatures at two equally-spaced points on opposite sides:

$$R(\theta,\varphi) = (\Delta T \star B)(\theta,\varphi) - \tfrac{1}{2}\left((\Delta T \star B)(\theta,\varphi+\alpha) + (\Delta T \star B)(\theta,\varphi-\alpha)\right), \tag{2.42}$$

where B denotes the beam pattern, as above, and α is the beam separation. For the Tenerife experiment, B is approximately a Gaussian of width $\sigma = 2^\circ\!.4$, and $\alpha \sin\theta = 8^\circ\!.1$. We have chosen a spherical coordinate system in which the beams are separated in azimuth φ. Once again, we can expand $\Delta T \star B$ in spherical harmonics and find that

$$R(\theta,\varphi) = \sum_{l,m} B_l a'_{lm} Y'_{lm}(\theta,\varphi)\left(1 - \tfrac{1}{2}(e^{im\alpha} + e^{-im\alpha})\right), \tag{2.43}$$

so that the window function is

$$W'_{lm} = B_l(1 - \cos m\alpha). \tag{2.44}$$

This expression is also valid for the real-valued spherical harmonics. Note that $W_{lm} \propto m^2$ for low $|m|$. Modes with low $|m|$ are therefore even more strongly suppressed than in the case of the MAX experiment. This is not surprising, of course, as a triple-beam experiment like Tenerife is designed to measure the second derivative $\partial^2\Delta T/\partial\varphi^2$, which is very small for modes with low $|m|$.

We have computed window functions that allow one to compute the response R in terms of ΔT. For many purposes it is convenient to talk about window functions for the mean-square power $\langle R^2\rangle$ instead of R. Specifically, one defines a window function $\mathcal{W}$ such that

$$\langle R^2\rangle = \frac{1}{4\pi}\sum (2l+1) C_l \mathcal{W}_l. \tag{2.45}$$

$\mathcal{W}_l$ is related to W_{lm} in a simple way. The instrument response is

$$R(\theta,\varphi) = \sum_{l,m} W_{lm} a_{lm} Y_{lm}(\theta,\varphi). \tag{2.46}$$

Squaring and taking an ensemble average, we have

$$\langle R^2(\theta,\varphi)\rangle = \sum_{l,m} C_l W_{lm} Y_{lm}^2(\theta,\varphi). \tag{2.47}$$

Therefore,

$$\mathcal{W}_l = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} W_{lm}^2 Y_{lm}^2(\theta,\varphi). \tag{2.48}$$

Note that in the case of an isotropic experiment like the DMR, in which W_{lm} is independent of m, the addition theorem allows us to simplify this to

$$\mathcal{W}_l = W_l^2. \tag{2.49}$$

2.5 The COBE DMR Data

The COBE DMR experiment has been described in detail elsewhere [11, 13, 18, 94, 105, 116, 161, 163]. In this section we will content ourselves with a brief summary of the relevant facts about the experiment and the data. The DMR experiment, which was one of three instruments aboard the COBE satellite, consisted of six differential radiometers with a beam size of 7° (FWHM) designed to measure intensity differences between pairs of points on the sky separated by 60°. There were two independent channels at each of three frequencies, 31, 53, and 90 GHz. The DMR recorded a highly redundant network of 60° temperature differences, which were inverted to produce sky maps containing an antenna temperature measurement for each of 6144 equal-area pixels. Each pixel contains noise, which appears to be Gaussian and has well-estimated variances and covariances. Separate maps are made for each of the six channels. Sky maps have been released for the first year and first two years of data; the full four-year data set is currently being analyzed and is expected to be released in early 1996.

The time-series data have been reduced to sky maps in both Galactic and ecliptic coordinate frames. The procedure used to make the maps was identical, but the locations of the pixels are different in the two cases. We have performed our analysis on the ecliptic maps, as these were the first to be publicly released. However, we have performed some tests on the Galactic maps as well, in order to assess the systematic error intrinsic to the map-making process. These tests are summarized in Section 4.5.

The six two-year sky maps are shown in Figure 2.1. The maps are Aitoff-projected in Galactic coordinates (although the maps were made from the ecliptic-projected data sets).

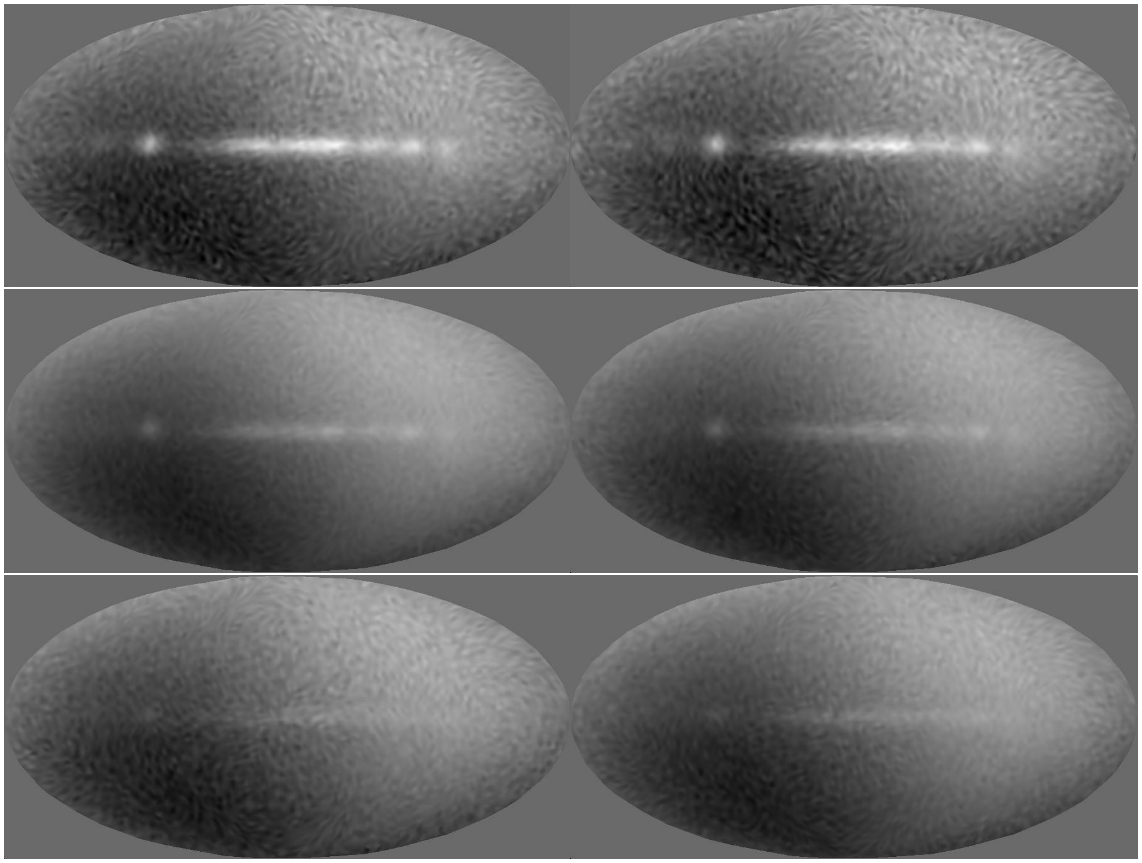

Figure 2.1: The COBE DMR data

The six two-year ecliptic DMR sky maps, shown in Aitoff projection in Galactic coordinates. The Galactic center $(l, b) = (0°, 0°)$ is at the center, and Galactic longitude increases from right to left. The top row shows the 31 GHz A and B maps, and the next two rows show the 53 and 90 GHz maps. In all maps, the greyscale ranges from -5 mK to 7 mK.

The Galactic center $(l, b) = (0^\circ, 0^\circ)$ is at the center, and Galactic longitude b increases from right to left across the map. We will make frequent use of Aitoff-projected all-sky maps in this dissertation. Details of the Aitoff projection are presented in Appendix A. The maps are dominated by two features: Galactic emission, which runs across the equator and appears most prominently in the low-frequency maps, and a dipole, which is presumed to be caused by the motion of the solar system with respect to the CMB rest frame.

Both of these features must be removed before we can see the underlying cosmic signal. Attempts have been made to remove the Galactic contribution using the fact that Galactic emission has a different spectral signature from CMB anisotropy [12, 13, 163]; however, this can only be done at the cost of significantly increasing the noise. Furthermore, it is not possible to separate the cosmic and Galactic signals perfectly, and the residual Galactic contamination is difficult to model. Since the Galactic contribution is concentrated near the Galactic plane, the most reliable way to remove this contaminant is simply to remove all pixels at low Galactic latitude. We will generally remove all pixels with $|b| < 20^\circ$, reducing the number of pixels in the map from 6144 to 4038.[12]

After excising the contaminated pixels, we need to remove the dipole from the data. Actually, the anisotropy caused in the CMB due to our motion at constant speed with respect to the CMB rest frame is almost, but not exactly a dipole. The kinematic anisotropy seen by an observer moving at velocity v is given by the full relativistic Doppler formula

$$1 + \frac{\delta T}{T} = \frac{\sqrt{1 - v^2/c^2}}{1 - (v/c)\cos\theta}, \tag{2.50}$$

where θ is the angle between the direction of motion and the direction of observation. If $v \ll c$, then this reduces to the usual nonrelativistic equation

$$\Delta T/T \approx (v/c)\cos\theta, \tag{2.51}$$

which is a pure dipole pattern. However, in our case, the second-order correction to this formula is not negligible. $v/c \sim 10^{-3}$, so the second order corrections have characteristic magnitude 10^{-6}. We shall see that the cosmic signal is at about the 10^{-5} level, so the second-order contribution gives a small but non-negligible correction to the cosmic signal. Expanding equation (2.50) to second order in v/c gives

$$\frac{\Delta T}{T} = (v/c)\cos\theta + (v/c)^2(\cos^2\theta - \tfrac{1}{2}). \tag{2.52}$$

---

[12] In maps that are pixelized in Galactic rather than ecliptic coordinates, the Galactic cut leaves 4016 pixels rather than 4038.

The first term is simply the usual dipole. The second term contains monopole and quadrupole contributions. Since the DMR is a differencing experiment, the monopole in the maps is of no significance in any case. However, it is important to remove the kinematic quadrupole. Since v/c is small, we can proceed perturbatively. We estimate the dipole by least-squares fitting using the simple nonrelativistic formula (2.51). Then once we know v, we can remove the kinematic quadrupole using equation (2.52).

Once we have removed the dipole, we often find it useful to average the maps together to reduce the noise. We perform the usual variance-minimizing weighted average in each pixel of the map. Let $\Delta T_{i\alpha}$ be the measured temperature anisotropy in pixel i of map α, and let $\sigma_{i\alpha}^2$ be the variance of the noise in that pixel. Then we average maps together with weights inversely proportional to the noise variances:

$$\Delta T_i^{(\mathrm{avge})} = \frac{\sum_\alpha \Delta T_{i\alpha}/\sigma_{i\alpha}^2}{\sum_\alpha 1/\sigma_{i\alpha}^2}. \tag{2.53}$$

The noise in the various maps is independent, so the noise averages in quadrature:

$$\sigma_i^{(\mathrm{avge})} = \left(\sum_\alpha 1/\sigma_{i\alpha}^2\right)^{-1/2}. \tag{2.54}$$

For the likelihood analyses with which we will be primarily concerned, this average map contains just as much information as a joint likelihood analysis of all of the maps. See Appendix B for a proof of this.[13]

Figure 2.2 shows the result of averaging together the two 53 GHz and the two 90 GHz maps in this way. Most of the analysis performed in later chapters will be on this average map. We do not include the 31 GHz maps in our final analysis, because they are more susceptible to Galactic contamination. In any case, these maps have significantly higher noise levels than the higher-frequency maps, and so we lose little information by excluding them. In fact, we have performed portions of the likelihood analysis described in chapters 4 and 5 on maps including and excluding the 31 GHz maps, with negligible differences in the final results. The one-sigma noise levels in the map shown in Figure 2.2 range from 56 μK to 108 μK. Because of variable sky-coverage due to the orbit of the COBE satellite, the noise levels vary across the sky in the manner shown in Figure 2.3.

Figure 2.4 is designed to illustrate the relative magnitudes of the signal and noise in the DMR data. The top panel shows a map of elevation on the Earth, projected in

---

[13] Strictly speaking, this statement is only true if the noise is uncorrelated from pixel to pixel. For the case of the DMR data, this condition is very nearly met [117].

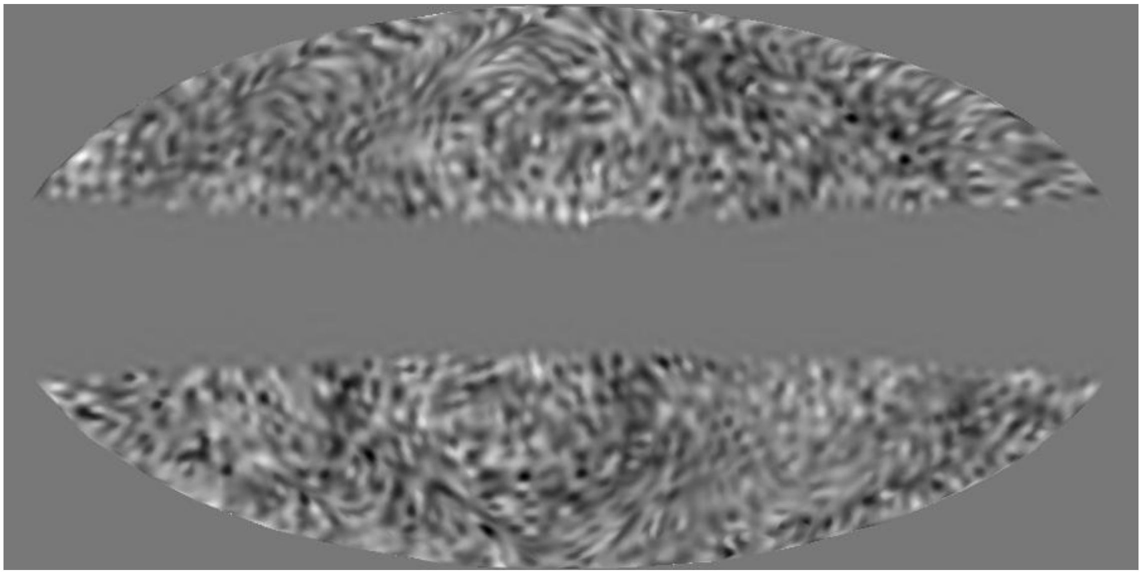

Figure 2.2: The galaxy-cut, dipole-subtracted, weighted-averaged DMR data

This map was made by averaging together the four maps at 53 GHz and 90 GHz after removing all pixels within 20° of the Galactic plane and fitting out a dipole. The projection is the same as in figure 2.1.

the same manner as the DMR maps.[14] If we take this data set and convolve it with a 7° FWHM beam, the result is the map in the center panel. Not surprisingly, small-scale features like the Andes are obscured by this smoothing. Finally, in the lower panel we have added noise to the data with the same characteristics as the DMR noise. (To be specific, we added independent Gaussian noise to each pixel, with amplitude proportional to the amplitude in the weighted-average two-year DMR map shown in Figure 2.2 and the constant of proportionality chosen to match the overall signal-to-noise ratio in that map.) The noise is quite effective in obscuring all but the most prominent large features. The figure also shows the result of applying a Wiener filter to the noisy world map, in an attempt to reconstruct the true underlying signal. We will describe this filter in detail in Chapter 6.

---

[14] Actually, the projection used is the mirror image of that used for the DMR maps, since sky maps look backwards when compared with Earth maps. We are outside the Earth, but inside the celestial sphere.

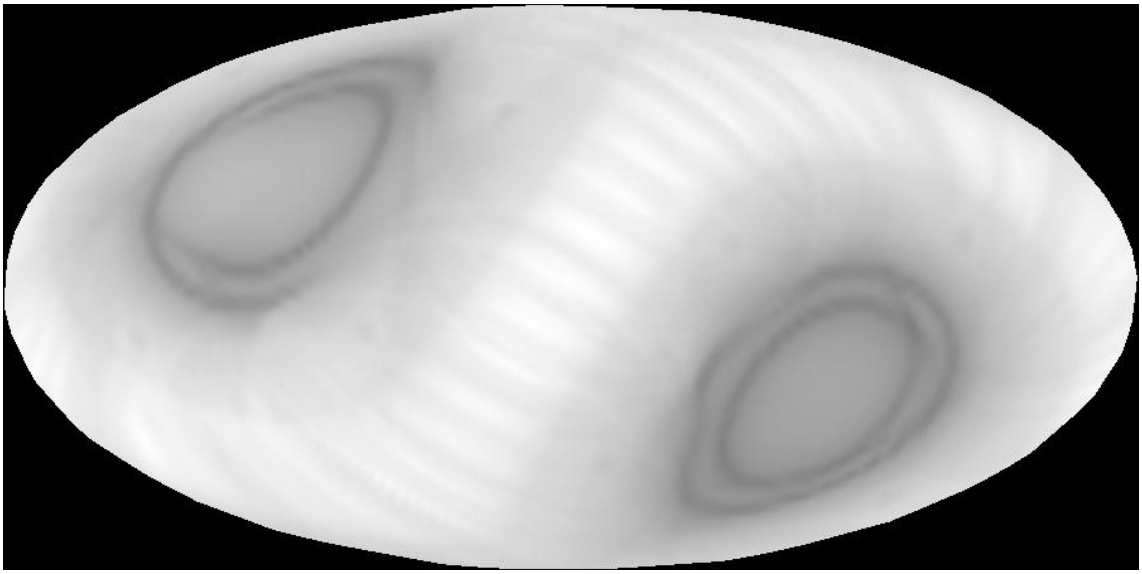

Figure 2.3: Noise levels in the COBE DMR data

The noise level in the two-year DMR sky map shown in Figure 2.2. The North and South celestial poles were scanned more frequently than other points, and therefore have lower noise levels. Due to the DMR 60° differencing strategy, points 60° from these points also have low noise; this is the reason for the two rings visible in the Figure.

2.6 Notation

Throughout this dissertation, we will make frequent use of techniques and results from linear algebra. In this section we will establish consistent notation for the various vectors and matrices we will have occasion to use.

Suppose we have a data set consisting of $\mathcal{N}$ pixels. (Typically, the data set will be the averaged two-year map in Figure 2.2 and $\mathcal{N}$ will be 4038.) Let d_i be the measurement corresponding to the ith pixel, and let $\vec{d}$ be the $\mathcal{N}$-dimensional vector $(d_1, \ldots, d_{\mathcal{N}})$. (Throughout this dissertation, we will denote vectors in real three-dimensional space by boldface type, and vectors in other spaces such as "pixel space" by arrows.) The datum d_i contains contributions from both signal and noise:

$$d_i = (\Delta T \star W)(\hat{\mathbf{r}}_i) + n_i. \tag{2.55}$$

Here ΔT is the true temperature anisotropy, $W(\hat{\mathbf{r}}, \hat{\mathbf{r}}')$ represents the window function of

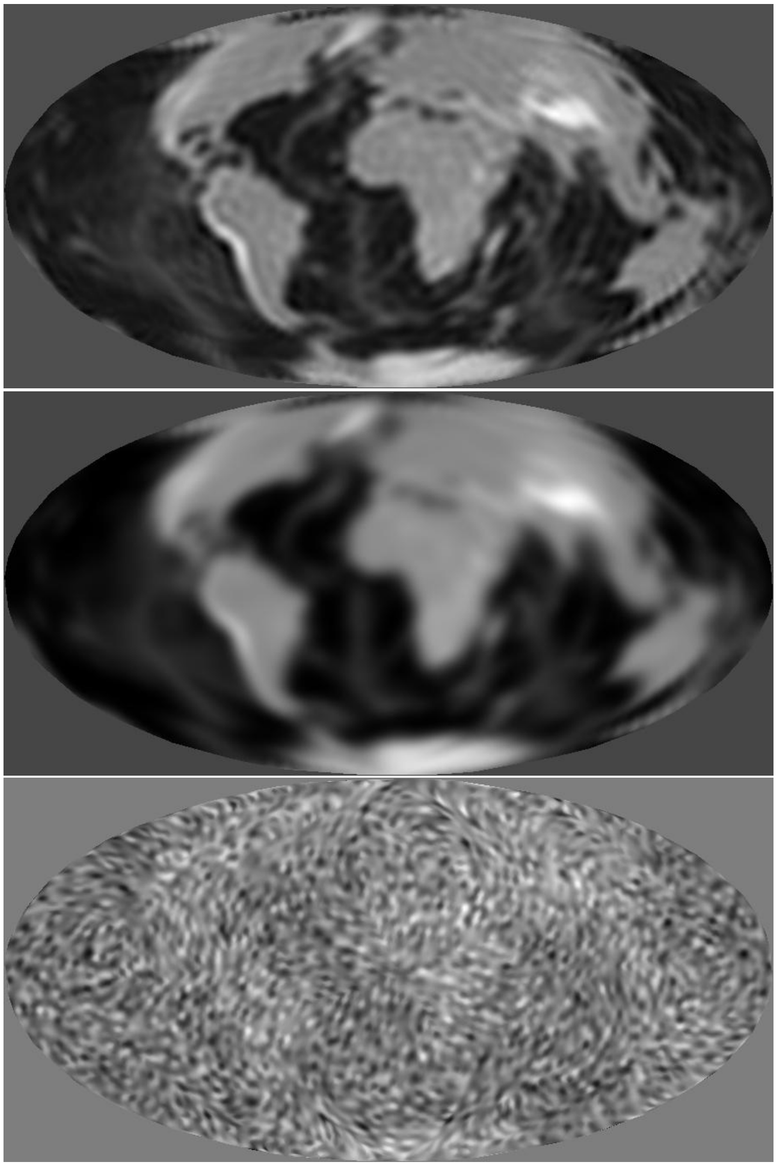

Figure 2.4: The world as it would be seen by COBE

The top panel shows a map of the elevation of the Earth's surface, projected in the same manner as Figure 2.1. The center panel shows the result of smoothing this map with the COBE beam pattern, and the third shows the result of adding noise with the same amplitude as the noise in the map of Figure 2.2. The fourth panel (on the next page) shows the result of applying a Wiener filter to the noisy data, as described in Chapter 6. Before applying the filter, all pixels within 20° of the equator were removed.

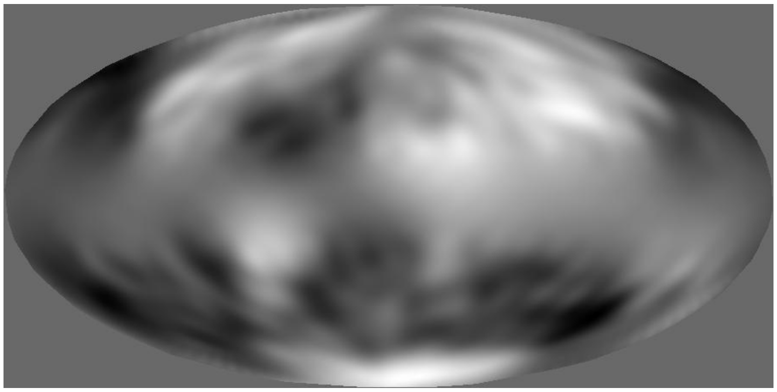

Figure 2.4 (continued)

the experiment, and the star denotes convolution. $\hat{\mathbf{r}}_i$ is the location on the sky of the ith pixel. n_i is a random variable representing the noise. In general, we will assume that n_i is Gaussian and has zero mean. We will denote the noise covariance matrix by $\mathbf{N}$:

$$N_{ij} = \langle n_i n_j \rangle . \tag{2.56}$$

For the COBE DMR data, noise correlations are very weak, and so $\mathbf{N}$ is very nearly diagonal [117]. (Throughout this dissertation, matrices will be denoted with capital boldface letters.)

If we expand ΔT in spherical harmonics as in equation (2.5), then we have

$$d_i = \sum_{l,m} W_l a_{lm} Y_{lm}(\hat{\mathbf{r}}_i) + n_i, \tag{2.57}$$

where W_l is the Legendre polynomial expansion of the window function, as described in Section 2.4.

Let us introduce some simplifying notation. We will denote a pair of indices (lm) by a single Greek index such as μ. The correspondence between the two is

$$\mu = l(l+1) + m + 1 \tag{2.58}$$

so that μ ranges from 1 to ∞ as l and m vary over their entire allowed ranges. If we introduce an $\mathcal{N} \times \infty$ matrix $\mathbf{Y}$ whose elements are

$$Y_{i\mu} = Y_\mu(\hat{\mathbf{r}}_i), \tag{2.59}$$

then equation (2.5) becomes simply

$$\vec{\Delta T} = \mathbf{Y} \cdot \vec{a}. \tag{2.60}$$

The most natural way to represent the window function is by a diagonal matrix $W_{\mu\nu} = W_l\delta_{\mu\nu}$, where l is the index corresponding to μ. Then equation (2.57) becomes

$$\vec{d} = \mathbf{Y}\cdot\mathbf{W}\cdot\vec{a} + \vec{n}. \tag{2.61}$$

(Here $\vec{a}$ and $\vec{n}$ are the vectors whose components are a_μ and n_i respectively.)

The covariance matrix of $\vec{a}$ is diagonal, and its nonzero elements are those of the angular power spectrum C_l:

$$C_{\mu\nu} \equiv \langle a_\mu a_\nu\rangle = C_l\delta_{\mu\nu}, \tag{2.62}$$

where l is the index corresponding to μ as before. We know the covariance matrices of $\vec{a}$ and $\vec{n}$, and we assume that the signal and noise are uncorrelated: $\langle a_\mu n_i\rangle = 0$. We can therefore write down the covariance matrix of the data vector $\vec{d}$:

$$\mathbf{M} \equiv \langle\vec{d}\cdot\vec{d}^T\rangle = \mathbf{Y}\cdot\mathbf{W}\cdot\mathbf{C}\cdot\mathbf{W}\cdot\mathbf{Y}^T + \mathbf{N}. \tag{2.63}$$

The matrix $\mathbf{W}\cdot\mathbf{C}\cdot\mathbf{W}$ is the beam-smoothed angular power spectrum, which we will refer to from time to time as $\bar{\mathbf{C}}$.

Until this point, we have ignored the fact that a monopole and dipole have been removed from the data. If we had complete sky coverage and no noise, this would not be a problem: removing the monopole and dipole from the data would simply correspond to setting $a_{00} = a_{1m} = 0$. In real life things are not so simple. Our least-squares estimate of the monopole and dipole will not in general coincide with the true values.

Suppose we begin with a data vector $\vec{d}_0$, and we remove best-fitting monopole and dipole terms by least-squares fitting. Let $\mathbf{Z}$ denote the $\mathcal{N}\times 4$ matrix whose columns are the first four columns of $\mathbf{Y}$. (These are of course the four columns representing the monopole and dipole.) Then the procedure for least-squares fitting involves choosing the four numbers $\vec{z} = (z_1, z_2, z_3, z_4)$ that minimize

$$\Delta = |\vec{d}_0 - \mathbf{Z}\cdot\vec{z}|^2. \tag{2.64}$$

Setting $\partial\Delta/\partial z_i = 0$, we find that $\vec{z} = (\mathbf{Z}^T\cdot\mathbf{Z})^{-1}\cdot\mathbf{Z}^T\cdot\vec{d}_0$. Removing the monopole and dipole subtracts $\mathbf{Z}\cdot\vec{z}$ from the data, so our new data set is

$$\vec{d} = \mathbf{P}\cdot\vec{d}_0 = (1 - \mathbf{Z}\cdot(\mathbf{Z}^T\cdot\mathbf{Z})^{-1}\cdot\mathbf{Z}^T)\cdot\vec{d}_0. \tag{2.65}$$

$\mathbf{P}$ is the projection operator that removes the monopole and dipole.

The covariance matrix in equation (2.63) actually applies to $\vec{d}_0$, not to $\vec{d}$. It must therefore be modified as follows:

$$\mathbf{M} = \mathbf{P}\cdot\mathbf{Y}\cdot\mathbf{W}\cdot\mathbf{C}\cdot\mathbf{Y}^T\cdot\mathbf{P}^T + \mathbf{P}\cdot\mathbf{N}\cdot\mathbf{P}^T \tag{2.66}$$

$$= \tilde{\mathbf{Y}}\cdot\bar{\mathbf{C}}\cdot\tilde{\mathbf{Y}}^T + \tilde{\mathbf{N}}, \tag{2.67}$$

where $\tilde{\mathbf{Y}} = \mathbf{P}\cdot\mathbf{Y}$ and $\tilde{\mathbf{N}} = \mathbf{P}\cdot\mathbf{N}\cdot\mathbf{P}^T$.

It is often a good approximation to set $\tilde{\mathbf{N}} = \mathbf{N}$. The reason is that the projection operator $\mathbf{P}$ couples most strongly to modes with large angular scales, and most of the noise power is on small scales. On the other hand, $\tilde{Y}_{i\mu}$ can differ quite significantly from $Y_{i\mu}$, especially for low values of μ. (Indeed, for $\mu \le 4$, $\tilde{Y}_{i\mu}$ vanishes.) The difference between $\mathbf{Y}$ and $\tilde{\mathbf{Y}}$ is a measure of the degree to which the spherical harmonics fail to be orthogonal over the cut sky.

2.7 A Polemic Regarding the COBE Quadrupole

We have already seen that CMB experiments in general, and the DMR in particular, have no sensitivity to the monopole component of the anisotropy and that the dipole is impossible to distinguish from the kinematic dipole caused by our motion with respect to the CMB center-of-momentum frame. We therefore include only modes with $l \ge 2$ in our data analysis, taking proper account (*e.g.*, by marginalizing) of our ignorance of the other modes. Some analyses of the DMR data have chosen to discard the quadrupole information in the data as well, on the grounds that it may be contaminated by systematic effects. In this section we will argue against this practice.

The COBE quadrupole appears to be a little bit too low to fit comfortably into popular theories. In particular, it may be too low to be consistent with a Harrison-Zel'dovich power spectrum. This is, of course, no reason to throw it out. On the contrary, there are few statistical *faux pas* greater than throwing out data points because they do not agree with our favorite theory. We must find independent justification before we can reasonably decide to ignore the quadrupole. Furthermore, since the decision to ignore the quadrupole appears to have been made after it was already known to be anomalously low, we should be especially wary of the dangers of biased data editing and should apply an especially high burden of proof before deciding to ignore it.

In fact, as we will argue in more detail below, the very fact that the COBE quadrupole is low argues *against* the hypothesis that it is contaminated. Since the quadrupole amplitude is a root-mean-square quantity, any contaminant in the data will tend to bias the quadrupole upwards. In fact, we will prove below that any theory that is ruled out by the COBE data due to the anomalously low quadrupole is ruled out even more strongly if the quadrupole is contaminated.[15]

We will divide this section, like Gaul, into three parts [35]. First, we will assess quantitatively whether the COBE quadrupole is anomalously low compared to a prediction based on a Harrison-Zel'dovich spectrum. Second, we will discuss whether there is any independent evidence, either within the data or from elsewhere, to suggest contamination of the quadrupole. Finally, we will argue that even if the quadrupole is contaminated, it still contains some useful information, and will suggest a procedure for accounting for this information.

2.7.1 Is the quadrupole low?

Bennett *et al.* [13] estimate the quadrupole in the two-year DMR data to be $\hat{Q} = 6 \pm 3\,\mu\mathrm{K}$, where $\hat{Q}$ is supposed to be an estimator of the quadrupole Q defined by

$$Q = \sqrt{\frac{1}{4\pi} \sum_{m=-2}^{2} a_{2m}^2}. \tag{2.68}$$

On the other hand, the ensemble-average quadrupole of a Harrison-Zel'dovich power spectrum normalized to COBE is about $21\,\mu\mathrm{K}$. There thus appears to be a *prima facie* discrepancy. However, Q^2 is chi-squared distributed with only five degrees of freedom, and therefore has a very large cosmic variance. We must therefore check to see how unlikely such a low quadrupole is.

The probability that the actual Q would come out as low as $6\,\mu\mathrm{K}$ is simply the probability that a random variable with a χ^2_5 distribution will take a value $\leq (6/21)^2$ times its expectation value. This works out to be 0.49%, although if we increase the bound by one standard error to $9\,\mu\mathrm{K}$, the consistency probability rises to 3.1%. We can make a slightly

---

[15] Strictly speaking, this statement is true only if the contaminant is statistically independent of the cosmic signal. The only way in which this assumption could be false is if an attempt had been made to remove contaminants in the data in a way that made use of the information in the data itself. We shall see below that the very low quadrupole estimate in [13] may indeed be the result of the process by which Galactic emission was removed from the data. The author is tempted to remark that anybody who processes his data in this way and makes no attempt to account for these effects deserves all he gets.

more realistic estimate of the consistency probability by assuming that $\hat{Q}$ is a measurement of Q convolved with Gaussian noise of standard deviation $\sigma = 3\,\mu\mathrm{K}$. The probability of getting a value as low as $6\,\mu\mathrm{K}$ out of this convolved probability distribution is 1.6%. The claim that the COBE quadrupole is too low to be consistent with a Harrison-Zel'dovich spectrum therefore seems initially to have some validity.

One should not take this sort of analysis too seriously, however, because $\hat{Q}$ is not a simple estimate of the quantity Q in equation (2.68). The data set used to make the estimate is a linear combination of the three maps of different frequencies, with coefficients chosen to minimize Galactic emission. The coefficients were chosen based in part on the signal in the sky maps themselves, and so presumably this process removed some of the noise contribution to the quadrupole as well as the Galactic contamination. This effect is difficult to quantify.

In addition, the estimate comes from a data set that does not cover the complete sky (although it does have only a 10° Galactic cut rather than the usual 20°), and as we shall describe in some detail in chapter 3, there is no way to estimate individual coefficients a_{lm} from such a data set without contamination from other coefficients. Therefore, even if we ignore noise, the probability distribution of the Bennett *et al.* estimate $\hat{Q}$ is therefore not the simple χ_5^2 distribution corresponding to the true quadrupole Q; its distribution is something more complicated, which would depend on all of the C_l's rather than just C_2.

These effects make the true probability distribution of the above quadrupole estimator quite difficult to model. We shall not attempt to do so here. Instead, we will estimate the quadrupole directly from the publicly-available two-year sky maps. Our estimator will still have contamination from other multipoles, but if we are willing to assume a particular power spectrum, we can still compute its probability distribution. Furthermore, since we have not tried to do anything fancy to remove Galactic contamination, we do not have to worry about having thrown away some unknown portion of the cosmic signal. To be sure, we do have to worry about residual Galactic contamination. But that is a contaminant that is statistically independent of the signal, and as we shall see below, such a contaminant always tends to bias the quadrupole upwards. By ignoring this possibility, we run the risk of erroneously concluding that the quadrupole is consistent with our hypothesis, but not of erroneously concluding that it is inconsistent.

We will take the simplest possible estimator of the quadrupole. We will use the two-year weighted-average map shown in Figure 2.2 and compute estimates of the coefficients a_{2m} for $|m| \leq 2$ by converting the integral in equation (2.6) into a discrete sum over pixels,

$$\hat{a}_{2m} = \frac{4\pi}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} d_i \tilde{Y}_{2m}(\hat{\mathbf{r}}_i), \tag{2.69}$$

or in the notation of Section 2.6,

$$\vec{a} = \frac{4\pi}{\mathcal{N}} \tilde{\mathbf{Y}}^T \cdot \vec{d} = \frac{4\pi}{\mathcal{N}} \mathbf{Y}^T \cdot \vec{d}. \tag{2.70}$$

These two expressions are equal because $\vec{d}$ has already had the monopole and dipole projected out of it. Since $\mathbf{P}$ is a projection operator, $\mathbf{P} \cdot \vec{d} = \vec{d}$, and so $\tilde{\mathbf{Y}} \cdot \vec{d} = \mathbf{Y} \cdot \mathbf{P} \cdot \vec{d} = \mathbf{Y} \cdot \vec{d}$.

We can then define a "quadrupole" estimator $\hat{Q}$ by

$$\hat{Q} = \sqrt{\frac{1}{4\pi} \sum_{m=-2}^{2} \hat{a}_{2m}^2}. \tag{2.71}$$

We could have chosen any of a number of other estimators. We could choose to weight the $\hat{a}_{2m}$'s differently based on their noise levels or their expectation values, or we could adopt a fancier procedure such as that of Tegmark [173], who has devised a set of quadrupole estimators that contain the minimum amount of contamination from other multipoles. We chose the estimator we did because it is simple and because it corresponds closely to what people mean when they talk about "the quadrupole" in the data.[16]

The estimator $\hat{Q}$ does not depend only on the true quadrupole Q: since we do not have complete sky coverage, that is impossible. In fact, it is contaminated by other multipoles and biased upwards by noise. However, it is more sensitive to the quadrupole than to any of the other multipoles. In any case, it has a probability distribution that we can calculate as long as we are willing to assume a particular form for the angular power spectrum C_l. We can therefore pose a statistically well-defined question that provides an assessment of whether the quadrupole in the DMR data is anomalously low, *viz.*:

> What is the probability of observing a value of $\hat{Q}$ as low as that observed in the real data, assuming a pure Harrison-Zel'dovich Sachs-Wolfe power spectrum with a normalization $Q_{\text{ps}} = 21\,\mu$K?

---

[16] We considered testing all of the various common estimators, but chose not to for fear of being accused of "fishing" for the estimator that best supports a particular conclusion. If we had used multiple estimators, we would have to decide at the end which one, or which combination, to use in drawing any final conclusions. There is no unbiased way to do that, and it is therefore safer to choose just one estimator, before computing anything, and stick with it.

m	$\hat{a}_{2m}$ (μK)	$\langle\hat{a}_{2m}^2\rangle^{1/2}$ (μK)	Probability
-2	2.22	22.6	0.078
-1	17.6	45.5	0.30
0	-9.02	26.4	0.27
1	29.0	45.3	0.48
2	15.6	22.7	0.51

Table 2.1: Quadrupole estimates

Column 2 contains the five quadrupole estimators defined by equation (2.69). Each is Gaussian with mean zero and standard deviation given in column 3. These standard deviations include both signal and noise contributions. The probability of getting a value as low as the observed value or lower is given in the final column.

To answer this question, we need to measure $\hat{Q}$ in the real data, and we need to compute its probability distribution based on the Harrison-Zel'dovich model. The first part is easy: all we need to do is compute the sum in equation (2.69). The values of the five estimators $\hat{a}_{2m}$ are given in Table 2.1. Adding together their squares, we get

$$\hat{Q} = 38.5\,\mu\text{K}. \tag{2.72}$$

At this point it may be prudent to remind the reader that we have made no attempt to remove noise bias from this estimator. The fact that it is much larger than the Bennett *et al.* estimator is therefore no cause for alarm. Since our goal is to compare this quantity with its theoretical probability distribution, it makes no difference whether we remove noise bias or not.

The quantities $\hat{a}_{2m}$ are Gaussian random variables of zero mean, so to specify their probability distribution (and therefore that of $\hat{Q}$), all we need is their covariance matrix. This matrix is simply

$$\langle\vec{a}\cdot\vec{a}^T\rangle = \left(\frac{4\pi}{\mathcal{N}}\right)^2 \mathbf{Y}^T\cdot(\tilde{\mathbf{Y}}\cdot\tilde{\mathbf{C}}\cdot\tilde{\mathbf{Y}}^T+\mathbf{N})\cdot\mathbf{Y}. \tag{2.73}$$

The square roots of the diagonal elements of this matrix are the standard deviations of the $\hat{a}_{2m}$'s, and are given in Table 2.1. The off-diagonal elements are generally small: the dimensionless ratio $\langle\hat{a}_{2m}\hat{a}_{2m'}\rangle/\sqrt{\langle\hat{a}_{2m}^2\rangle\langle\hat{a}_{2m'}^2\rangle}$ is always less than 2% in absolute value when $m\neq m'$.

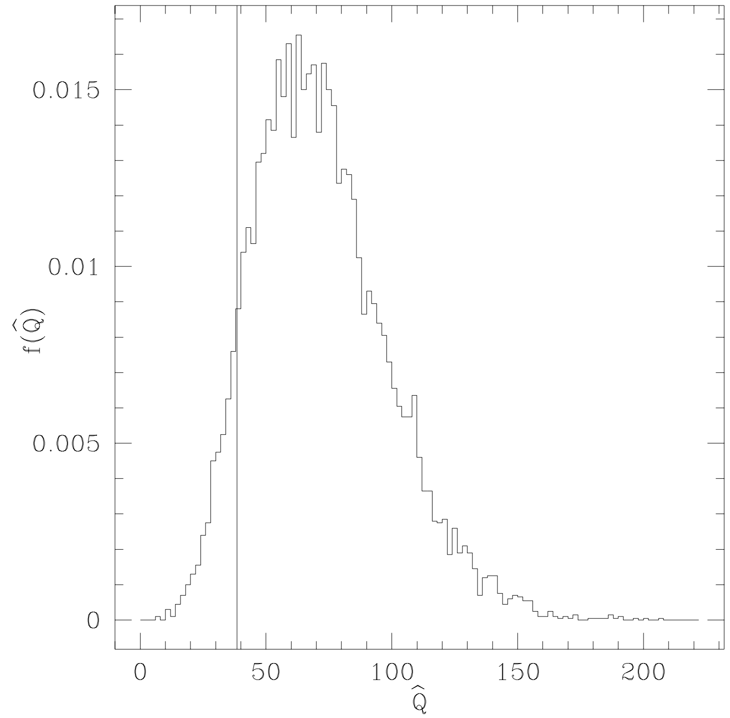

Figure 2.5: The probability distribution of $\hat{Q}$

The probability distribution of the quadrupole estimator $\hat{Q}$ as estimated from 10000 random realizations assuming a Harrison-Zel'dovich Sachs-Wolfe spectrum with a normalization $Q_{\text{ps}} = 21\,\mu\text{K}$. The value of $\hat{Q}$ from the real data is shown as a vertical line.

The probability distribution of $\hat{Q}$ is not easy to compute analytically, but we can estimate it using Monte Carlo techniques. We know the joint probability distribution of the five $\hat{a}_{2m}$'s. We draw random elements from this distribution and compute $\hat{Q}$ from these random numbers. Figure 2.5 shows the probability distribution of $\hat{Q}$ estimated from 10000 such trials. The value obtained from the real data is marked with a vertical line. It is clear that although $\hat{Q}$ is lower than its estimated expected value, it is not unreasonably low. In fact, 8.2% of our realizations gave lower values of $\hat{Q}$ than the real data, so we are unable to conclude that the quadrupole is anomalously low compared to a Harrison-Zel'dovich power spectrum at better than 92% confidence.

2.7.2 Is the quadrupole contaminated?

On the basis of this analysis, we can conclude that there is no strong evidence that the two-year DMR quadrupole is inconsistent with a simple Harrison-Zel'dovich model, although it is somewhat lower than the mean value it would be expected to take in such a model. In this section we will argue that the fact that the quadrupole is somewhat lower than its expectation value provides a fairly strong argument *against* the hypothesis that it is contaminated. We will demonstrate this in the following way. Let q be an estimator of the quadrupole that takes the value q_0 for the actual data. (q could be our estimator $\hat{Q}$ from the previous section, or it could be something else.) When we say that q is anomalously low, we mean that the probability $P(q < q_0)$ is a small number. We will show below under fairly general assumptions that adding a contaminant to the data always lowers $P(q < q_0)$. In other words, the hypothesis of contamination always worsens the fit to the data.

Assume that the pixel vector $\vec{d}$ is the sum of a cosmic signal, noise, and a contaminant vector,

$$\vec{d} = \Delta\vec{T} + \vec{n} + \vec{z}, \tag{2.74}$$

where $\vec{z}$ is the contaminant. The contaminant need not be Gaussian; the only assumption we will make is that it is independent of the cosmic signal. As we noted above, this may not be true of the Bennett *et al.* analysis, in which Galactic modeling was performed on the data. It should be true of the data set we are using, however, in which no Galactic modeling has been done. Let us also assume that our quadrupole estimator q is a root-mean-square quantity, so that q^2 is quadratic in $\vec{d}$,

$$q^2 = \vec{d}^T \cdot \mathbf{A} \cdot \vec{d} \tag{2.75}$$

for some symmetric $\mathcal{N} \times \mathcal{N}$ matrix $\mathbf{A}$.[17] Since the quadrupole is a positive quantity, it is reasonable to take $\mathbf{A}$ to be nonnegative definite. This means that we can find $k \leq \mathcal{N}$ eigenvectors $\vec{e}_i$ with associated eigenvalues λ_i and write

$$\mathbf{A} = \sum_{i=1}^{K} \lambda_i \vec{e}_i \cdot \vec{e}_i^T. \tag{2.76}$$

Let us consider first the case $k = 1$. Then

$$q^2 = \lambda_1 (\vec{e}_1 \cdot \vec{d})^2 = (q_* + \zeta)^2, \tag{2.77}$$

---

[17] People generally subtract some constant from a statistic like this in order to remove noise bias. This shifts q^2 and q_0^2 by a constant amount and therefore has no effect on the consistency probability $P(q < q_0)$.

where $q_* = \lambda_1^{1/2}\vec{e}_1 \cdot (\vec{\Delta T} + \vec{n})$ is the value that q would take in the absence of a contaminant $\vec{z}$ and $\zeta = \lambda_1^{1/2}\vec{e}_1 \cdot \vec{z}$. q_* has a Gaussian probability density f_{q_*}, and ζ has a probability density f_ζ about which we know nothing. We do know that ζ and q_* are independently distributed, though, so the joint probability density of q_* and ζ is the product $f_{q_*} f_\zeta$. The consistency probability is therefore

$$P(q < q_0) = \int_0^{q_0} dq_*\, f_{q_*}(q_*) \int_0^{\sqrt{q_0^2 - q_*^2}} d\zeta\, f_\zeta(\zeta). \tag{2.78}$$

The ζ integral in the above expression is always less than or equal to 1 (since f_ζ is a probability distribution), and is equal to 1 only if f_ζ has support only in the interval $|\zeta| \le \sqrt{q_0^2 - q_*^2}$. We therefore have the inequality

$$P(q < q_0) \le \int_0^{q_0} dq_*\, f_{q_*}(q_*) = P(q_* < q_0), \tag{2.79}$$

and equality holds only if f_ζ has support only at $\zeta = 0$, or in other words only if there is no contaminant at all. The right-hand side of this expression is simply the probability of getting as low a value of q as we observe under the hypothesis that there is no contaminant. We have therefore shown that adding a contaminant necessarily reduces the consistency probability $P(q < q_0)$.

The case where k, the rank of $\mathbf{A}$, is greater than 1 can be handled in a similar way. Define k-dimensional vectors $\vec{q}_*$ and $\vec{\zeta}$ with components $q_{*i} = \lambda_i^{1/2}\vec{e}_i \cdot (\vec{\Delta T} + \vec{n})$ and $\zeta_i = \lambda_i^{1/2}\vec{e}_i \cdot \vec{z}$. Then

$$q^2 = |\vec{q}_* + \vec{\zeta}|^2. \tag{2.80}$$

Furthermore, $\vec{q}_*$ has a multivariate Gaussian distribution. We can replace the integrals in equation (2.78) with k-dimensional integrals over $\vec{q}_*$ and $\vec{\zeta}$, and the argument goes through exactly as before.

We saw in the previous section that the quadrupole estimate $\hat{Q}$ is a little bit low compared to our expectation (based on a flat Harrison-Zel'dovich spectrum). One might be tempted to argue that the reason for this is that there is something contaminating the data. In this section we have shown that this cannot be the case. A contaminated data set would be even less likely to produce the observed quadrupole than an uncontaminated data set. The hypothesis of a Harrison-Zel'dovich power spectrum without a contaminant has a consistency probability of 8%, and so is reasonably consistent with the data, but any significant contaminant would reduce this consistency probability to unacceptable levels.

For clarity, let us reiterate at this point that the phrase "the observed quadrupole" refers to a quadrupole estimate based on the Galaxy-cut data set, and is therefore something quite distinct from the true quadrupole. The relevant statistical question to ask is simply whether this estimate is low compared to its own theoretical probability distribution, and this question is logically independent of the question of how well the estimate reproduces the (unmeasurable) true quadrupole.

Of course, we can change the power spectrum to predict a lower quadrupole and thereby get a better consistency probability. Some open-Universe models do precisely that, and are therefore better fits to the DMR data than flat models [72]. A power spectrum with a low quadrupole would have more room to incorporate a contaminant, but since these models fit the data well, there is still no reason to suspect contamination. The bottom line is that we should suspect contamination only if some particular multipole moment is anomalously *high* compared to our theoretical expectation, not if one is anomalously low. Even in this case, of course, we had better have a strong independent reason to suspect contamination: it is never acceptable to edit out portions of one's data simply because they don't fit a particular theory.

2.7.3 What should we do about it?

Let us suppose we did have independent evidence to suggest a quadrupolar contaminant in the DMR data. What should we do about it? The simplest thing to do is to treat the quadrupole in the same way as the monopole and dipole and "throw it away." In practice, one generally throws away unwanted modes by marginalizing over them: for example, to marginalize over the monopole, we assume that all possible values of the unknown coefficient a_{00} are equally likely, and we replace the likelihood L with the integral $\int L\, da_{00}$ over all possible values of this coefficient. (We will discuss marginalization in more detail in Section 4.1.) This is a reasonable procedure to adopt for the monopole and dipole, for which we really do have no useful information; however, it is not correct to apply it to the quadrupole, which still contains some information even if it is contaminated. To take an extreme example, we can state with confidence that the DMR data rule out any theory which predicts that $C_2 = 10^{10}\,\mu\mathrm{K}^2$, since the observed quadrupole is some eight orders of magnitude smaller than this, and it is exceedingly improbable that such a low quadrupole would arise in this theory. Furthermore, this conclusion is valid even if we allow for the possibility that the quadrupole is contaminated: in fact, the result of the previous section is that the conclusion is strengthened if we assume contamination. If we threw out the quadrupole by marginalizing, we would lose the ability to rule out these theories.

This particular example is of course absurdly exaggerated, but the issue it raises is quite genuine. Consider the case of cosmological-constant cold dark matter models. As we shall see in section 5.4, these models are somewhat disfavored by the DMR data, in large part because the models predict large-scale power that is missing in the data. When we marginalize over the quadrupole as well as the monopole and dipole, the constraints on these models are greatly weakened. These constraints are weaker than they need to be, because we have thrown away too much information by marginalizing over the quadrupole.

We will now suggest an alternate procedure. Suppose that we suspect that our data are contaminated and are willing to assume that this contaminant only (or at least predominantly) corrupts the quadrupole. Then the angular power spectrum of the beam-smoothed data will no longer be simply $\bar{C}_l \equiv W_l^2 C_l$, but rather

$$\bar{C}_l^\star = \bar{C}_l + Z\delta_{l2}. \tag{2.81}$$

That is, C_2 will be increased by some amount Z and all other multipoles will stay the same.[18] Z is the variance of the contaminant, and so is nonnegative.

We are interested in testing hypotheses about $\bar{C}_l$. Let us adopt a Bayesian attitude, as we will throughout much of this dissertation. The natural thing to do is then to marginalize over the nuisance parameter Z. Specifically, if $L(\bar{C}_l)$ is the likelihood associated with the power spectrum $\bar{C}_l$, we should compute the marginal likelihood

$$L_{\rm marg}(\bar{C}_l) = \int_0^\infty dZ\, L(\bar{C}_l^\star) f(Z), \tag{2.82}$$

where $f(Z)$ is some prior distribution for Z. In the absence of any better ideas, we might choose $f(Z)$ to be uniform in Z (for $Z > 0$), or we might choose a prior that is uniform in the expected amplitude $Z^{1/2}$ of the contaminant by setting $f(Z) \propto Z^{-1/2}$.[19]

We will illustrate the effects of marginalizing over Z with a simple example. Suppose we had performed an experiment with complete sky coverage. Then we could estimate the

---

[18] The probability distribution of a_{2m} may no longer be Gaussian, but in the absence of a better model, let us assume that it is.

[19] Neither of these priors is normalizable. We are meant to imagine that they are valid over the range of Z of interest to us, not necessarily over all Z. We will see below that very large values of Z are always highly unlikely anyway, so it doesn't matter how we treat them. We must rule out another potentially attractive prior, namely one that is uniform in $\ln Z$, because it diverges for small Z where the likelihoods are large.

individual coefficients a_{lm}. Assuming homoscedasticity (*i.e.*, equal noise levels in all pixels), our estimates $\hat{a}_{lm}$ would have independent noise, which we take to be Gaussian. In such an experiment the random variables $\hat{a}_{lm}$ are uncorrelated, and so the likelihood $L(\bar{C}_l)$ factors into a contribution from the quadrupole C_2 and a contribution from all other modes:

$$L(\bar{C}_l) = L_Q L_{\rm other}. \tag{2.83}$$

We can therefore understand the effects of marginalizing over the quadrupole by considering only L_Q. If we assume no contamination, then L_Q is simply

$$L_Q = S_2^{-5/2} e^{-2\pi\hat{Q}^2/S_2}, \tag{2.84}$$

where $\hat{Q}$ is the quadrupole estimated from the data according to equation (2.71), $S_2 = \bar{C}_2 + \sigma_2^2$, and σ_2^2 is the noise variance in the quadrupole. We will be interested in treating the likelihood as a function of the theoretical quadrupole $\bar{C}_2$, so let us introduce a dimensionless variable

$$r = \frac{5S_2}{4\pi\hat{Q}^2}, \tag{2.85}$$

which is linear in $\bar{C}_2$.

If we hypothesize a contaminant, the likelihood changes to

$$L_Q = (S_2 + Z)^{-5/2} e^{-2\pi\hat{Q}^2/(S_2+Z)}. \tag{2.86}$$

Marginalizing over Z with a uniform prior, we get

$$\begin{aligned} L_{\rm marg} &= \int_0^\infty dZ\,(S_2 + Z)^{-5/2} e^{-2\pi\hat{Q}^2/(S_2+Z)} & (2.87)\\ &= (2\pi\hat{Q}^2)^{-3/2} \int_0^{5/2r} dy\, y^{1/2} e^{-y} & (2.88)\\ &= (2\pi\hat{Q}^2)^{-3/2} \left(\frac{\sqrt{\pi}}{2} - \Gamma\left(\frac{3}{2}, \frac{5}{2r}\right)\right). & (2.89) \end{aligned}$$

This function approaches a constant as $r \to 0$ and decays as $r^{-3/2}$ for large r.

If we adopt a prior that is uniform in $Z^{1/2}$, then

$$\begin{aligned} L_{\rm marg} &= \int_0^\infty dZ\,(S_2 + Z)^{-5/2} e^{-2\pi\hat{Q}^2/(S_2+Z)} Z^{-1/2} & (2.90)\\ &= \frac{25}{12\pi^2\hat{Q}^4 r^2}\ {}_1F_1\left(2, \frac{5}{2}, -\frac{5}{2r}\right), & (2.91) \end{aligned}$$

where ${}_1F_1$ is a hypergeometric function. This function also approaches a constant as $r \to 0$, but it decays as r^{-2} when r is large.

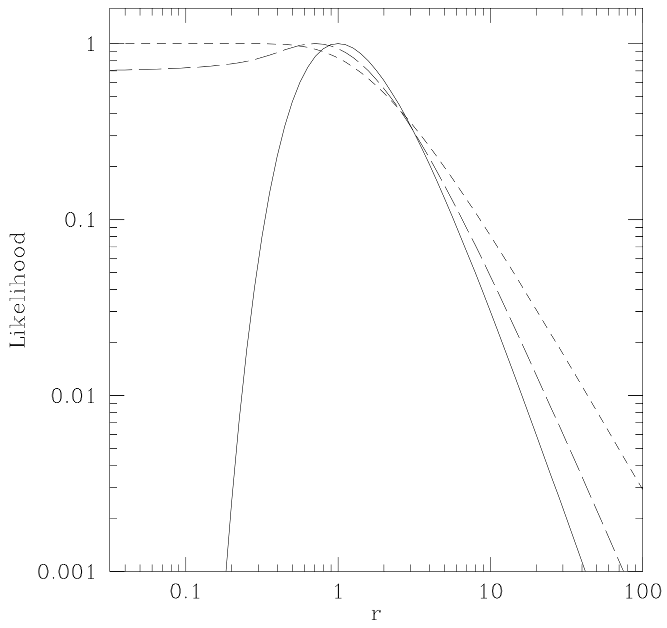

Figure 2.6: Marginal Likelihoods for the Quadrupole

The solid line shows the likelihood L_Q corresponding to the hypothesis of no contamination. The dashed lines show the two marginal likelihoods described in the text. The short-dashed line corresponds to a prior uniform in Z and the long-dashed line corresponds to a prior uniform in $Z^{1/2}$. All are arbitrarily normalized. The practice of throwing the quadrupole information away entirely would correspond to a horizontal line on this graph.

Both of these marginal likelihoods are plotted in Figure 2.6, along with the contaminant-free likelihood L_Q from equation (2.84). We can draw some important conclusions from this figure. First, the assumption of contamination eliminates our ability to eliminate models that have a very low predicted quadrupole. When no contaminant is assumed, the likelihood drops exponentially for low r (*i.e.*, low $\bar{C}_2$), but as soon as we hypothesize a contaminant, the marginal likelihood is flat for low r. This is not at all surprising, of course: if a contaminant is present, all theories with low $\bar{C}_2$ look identical, since the quadrupole anisotropy is dominated by the contaminant. The behavior of the likelihood at large r is more interesting. The fact that the likelihood still drops off rapidly reinforces and quantifies our earlier conclusion. If a model is ruled out because of a low quadrupole, hypothesizing a contaminant will not save it.

In contrast, the widespread practice of throwing away the quadrupole information entirely, *e.g.*, by marginalizing over the coefficients a_{2m}, corresponds to replacing L_Q by a constant. The fact that the likelihoods in Figure 2.6 are far from constant implies that this procedure throws away far more information than is necessary or appropriate.[20]

So what should we do with the COBE quadrupole? The best thing to do is nothing at all. Since there is no independent evidence for contamination, and since the rather low observed amplitude argues against the contamination hypothesis, we should proceed under the assumption that it is not contaminated. Even if, for some reason, we decide to hypothesize that there is a contaminant that only or primarily affects the quadrupole, we still should not throw away the quadrupole information entirely: even if the quadrupole is contaminated, the fact that it is observed to be low still provides evidence against theories that predict a large $\bar{C}_2$. One way to account for a contaminant without erroneously throwing away this information is to marginalize over Z as described above.

There are, of course, other procedures one could adopt. For example, if we are particularly worried about Galactic contamination, then we might want to consider throwing away only those modes that couple most strongly to the Galaxy. Since the Galaxy is fairly azimuthally symmetric at high latitudes and is roughly symmetric about the equator, we would worry most about modes with $m = 0$ and l even. We might therefore be inclined to throw away a_{20} and a_{40}, rather than the five quadrupole coefficients a_{2m}. However, Table 2.1 indicates that the low estimated quadrupole is not primarily due to a_{20}, and so throwing away a_{20} will not solve the "problem" of the low COBE quadrupole. Incidentally, likelihood-based techniques of the sort we will describe in Chapter 4 automatically give fairly low weight to modes like Y_{20} and Y_{40}, since these modes couple strongly to the monopole. We therefore expect that throwing away these modes would make little difference to the final results of such an analysis.

---

[20] To be fair, we should point out that the shapes of the likelihood curves in Figure 2.6 result from some combination of the quadrupole information in the data and the assumed prior. One might therefore worry that the shapes of these curves reflect primarily the extra information hidden in the prior, rather than anything in the data itself. However, the primary conclusion we wish to draw from Figure 2.6, that the likelihood is quite low for large values of r, appears to be true for any reasonably broad prior. This is not surprising: we are simply expressing the fact that high values of $\bar{C}_2$ are improbable whenever the observed quadrupole is low.

Chapter 3

Simple Models for Anisotropy Analysis

Far dearer is the honest simple friend,
Than one whose quicker wit is train'd to ill.
Euripides (410 B.C.)

3.1 Introduction

Before embarking on our quest to extract every drop of information from the actual COBE DMR data, it may be instructive to consider some toy models. In this Chapter, we will discuss three different ways one might try to estimate the spherical harmonic coefficients a_{lm} or constrain the theoretical power spectrum C_l from a hypothetical CMB anisotropy data set. This hypothetical data set is inspired by the COBE DMR data: it covers the whole sky with the exception of a strip of half-width 20° about the equator. However, we make no attempt in this chapter to simulate the DMR data in detail. We will spend enough time doing that in subsequent chapters. Our purpose in this chapter is to get a feeling for the problems caused by incomplete sky coverage. We will be interested in two quite distinct classes of questions:

1. How well can we estimate "local" quantities such as the coefficients a_{lm}?
2. How well can we constrain the power spectrum C_l?

In particular, it is clear that we cannot hope to measure any coefficient a_{lm} in the absence of complete sky coverage. This problem is not even well-posed. Nonetheless, people for some reason are interested in estimating the CMB quadrupole Q. We will see how well this task can be accomplished. If our goal is to constrain cosmological models, we are less interested in local quantities Q than in the power spectrum normalization $Q_{\rm ps}$. We will therefore see how well our various methods do at estimating power spectrum parameters. We will proceed under the assumption that the CMB anisotropy power spectrum is of the Sachs-Wolfe form given in equation (2.29), so that the two parameters n and $Q_{\rm ps}$ determine the power spectrum.

The first method is based on techniques developed by Peebles [135] and Hauser & Peebles [81]. We will therefore call it the Hauser-Peebles method. This method, which has been used in the past to analyze galaxy catalogues [152], is the most straightforward scheme for taking account of incomplete sky coverage. If one knew the function ΔT over the whole sky, then a_{lm} could be determined by integration:

$$a_{lm} = \int_S \Delta T(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}})\, d\Omega, \tag{3.1}$$

where $d\Omega$ is the element of solid angle corresponding to $\hat{\mathbf{r}}$ and the integral is taken over the entire sphere. With incomplete sky coverage, we can compute estimated coefficients b_{lm} by integrating over the observed region R:

$$b_{lm} = N_{lm} \int_R \tilde{\Delta T}(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}})\, d\Omega. \tag{3.2}$$

Here $\tilde{\Delta T}$ is the function obtained by subtracting the observed monopole and dipole components from ΔT. N_{lm} is a factor chosen to normalize b_{lm} appropriately. The simplest choice would be to set N_{lm} equal to the reciprocal of the fraction of sky covered; however, other methods are sometimes used.

The second method is least-squares fitting. Here one determines the estimated coefficients b_{lm} within some range $l_{min} \le l \le l_{max}$ so that the spherical harmonic expansion and the original function match as closely as possible over the observed region. The b_{lm}'s are chosen to minimize the quantity

$$\Delta \equiv \int_R \left(\Delta T(\hat{\mathbf{r}}) - \sum_{l=l_{min}}^{l_{max}} \sum_{m=-l}^{l} b_{lm} Y_{lm}(\hat{\mathbf{r}}) \right)^2 d\Omega. \tag{3.3}$$

(Since our goal in this chapter is to develop simple "toy models," we will assume that any noise in our experiment is uniform over the observed region R. We therefore adopt a uniform weight function in this expression.)

The final method is to compute the angular correlation function. Although this does not involve the spherical harmonics directly, the correlation function is a simple function of the coefficients a_{lm}:

$$C(\theta) = \frac{1}{4\pi} \sum_l A_l P_l(\cos\theta), \tag{3.4}$$

where P_l is a Legendre polynomial and A_l is the total power in multipole l:

$$A_l = \sum_{m=-l}^{l} |a_{lm}|^2 , \tag{3.5}$$

which implies that $\langle A_l \rangle = (2l+1)C_l$. The correlation function can be estimated with incomplete sky coverage, and equation (9) can be inverted to get the power C_l. One still needs to determine, however, whether the power derived in this way is a good estimate of the actual power.

We will find that all of these methods suffer great difficulties when used to determine properties of the CMB anisotropy as viewed from our particular vantage point; for example, none is able to reproduce accurately the local quadrupole. This is not surprising: after all, the quadrupole moment is not uniquely determined by an incomplete data set. The situation is somewhat different when the data are used to constrain cosmological models. For example, one might try to determine the normalization and slope of the power spectrum that underlies the observed fluctuations. This sort of analysis is less dramatically affected by masking; some methods are quite reliable even in the case of incomplete sky coverage. However, one must choose one's method of analysis with care, as some methods do suffer noticeable degradation as a result of masking.

3.2 Hauser-Peebles Method

Masking out part of the sky destroys the orthogonality of the spherical harmonics, and introduces correlations between the measured coefficients b_{lm} [135]. Let R be the observed region of the sky, and assume that R is azimuthally symmetric. (Note that in the analysis of real data, this assumption is unlikely to be strictly true. However, it is good enough for our present purpose.) The nonorthogonality of the spherical harmonics over the observed portion of the sky is characterized by the integrals

$$W_{ll'm} = \int_R Y_{lm}(\hat{\mathbf{r}}) Y_{l'm}(\hat{\mathbf{r}})\, d\Omega. \tag{3.6}$$

If R were the whole sphere, then $W_{ll'm}$ would be $\delta_{ll'}$. If R is symmetric about the equator, as in the case we are studying, then $W_{ll'm}$ is zero if $l - l'$ is odd. Subject to this constraint, $W_{ll'm}$ is largest when $l \sim l'$, and drops off rapidly when $|l - l'|$ is large. (If the region R is not azimuthally symmetric, then modes with different values of m are coupled to one another, and $W_{ll'm}$ must be replaced by $W_{ll'mm'}$.)

Before we compute the estimated coefficients, we need to subtract the observed monopole and dipole from ΔT. Suppose that the actual coefficients of the CMB anisotropy are a_{lm}. The monopole coefficient a_{00} is unmeasurable, and the coefficients a_{1m} are dominated by the kinematic dipole. Before we can extract any cosmological information from the higher multipoles, we must first subtract off these terms.

An observer who can only see a region R of the sky will measure the average value of ΔT to be

$$\overline{\Delta T} = \frac{1}{\Omega_R} \int_R \Delta T(\hat{\mathbf{r}})\, d\Omega = \frac{1}{\sqrt{4\pi} W_{000}} \sum_l a_{l0} W_{l00}, \tag{3.7}$$

and the observed value of the coefficient a_{00} is $a_{00}^{(observed)} = \sqrt{4\pi}\,\overline{\Delta T}$. If the observer then performs a least-squares fit to a dipole, the observed coefficients will be

$$a_{1m}^{(observed)} = \frac{1}{W_{11m}} \sum_l a_{lm} W_{l1m}. \tag{3.8}$$

Subtracting the monopole and dipole from the observations is equivalent to replacing the function ΔT by another function $\tilde{\Delta T}$, whose spherical harmonic coefficients are

$$\tilde{a}_{lm} = \begin{cases} a_{lm} - a_{lm}^{(observed)} & \text{if } l \le 1 \\ \\ a_{lm} & \text{if } l > 1. \end{cases} \tag{3.9}$$

Once the monopole and dipole have been subtracted, the estimated spherical harmonic coefficients are

$$\begin{aligned} b_{lm} &= N_{lm} \int_R \tilde{\Delta T}(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}})\, d\Omega & (3.10) \\ &= N_{lm} \int_R \sum_{l'} \tilde{a}_{l'm} Y_{l'm}(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}})\, d\Omega & (3.11) \\ &= N_{lm} \sum_{l'} \tilde{a}_{l'm} W_{l'lm}. & (3.12) \end{aligned}$$

It is easy to check by direct calculation that $b_{00} = b_{1m} = 0$, as expected.

How should one choose the normalization factors N_{lm}? There are different procedures that one might consider. The simplest is simply to set $N_{lm} = 1/f_R$, where f_R is the fraction of the sphere covered by R. However, a different normalization scheme has generally been used in the past in analyzing cosmological data [135]. The prescription is to set

$$N_{lm} = W_{llm}^{-\frac{1}{2}}. \tag{3.13}$$

The reason for choosing this factor is that the mean-square values of the estimated coefficients are the same as the actual mean-square values in the limit that C_l varies much more slowly than $W_{ll'm}$: roughly speaking, one can say that

$$\left\langle |b_{lm}|^2 \right\rangle \approx C_l \qquad \text{if} \qquad \frac{1}{C_l}\frac{dC_l}{dl} \ll \frac{1}{W_{ll'm}}\left.\frac{dW_{ll'm}}{dl'}\right|_{l=l'} \tag{3.14}$$

As we will see, this approximation is fairly good for large l in the case of a Harrison-Zel'dovich power spectrum. In any case, since this is the normalization scheme that has been most commonly used in the past, and since there is no obvious better choice, we have chosen to use it in this work.

Using equation (3.12), it is easy to check that $\langle b_{lm} \rangle = 0$ and to compute the correlations between different b_{lm}'s:

$$\langle b_{lm} b_{l'm'} \rangle = N_{lm} N_{l'm'} \sum_{l_1,l_2} W_{ll_1m} W_{l'l_2m'} \langle \tilde{a}_{l_1m} \tilde{a}_{l_2m'} \rangle \tag{3.15}$$

$$= \delta_{mm'} \frac{1}{W_{llm} W_{l'l'm}} \sum_{l''} W_{ll''m} W_{l'l''m} \tilde{C}_{l''}. \tag{3.16}$$

Note that the b_{lm}'s are not independently distributed: there are correlations between them.

To get a feel for what happens when one uses these coefficients in analyzing a COBE-like CMB experiment, we performed some Monte Carlo simulations. We assumed that the actual temperature fluctuation $\Delta T(\hat{\mathbf{r}})$ was drawn from a Gaussian random field with an $n = 1$ power spectrum. We chose coefficients a_{lm} for $2 \leq l \leq 30$ from Gaussian distributions with widths given by (5). For each of 10,000 realizations, we computed the estimated coefficients b_{lm} for $2 \leq l \leq 20$ from equation (3.12), assuming that the observed region R consisted of all points further than 20° from the equator. Figure 3.1 shows the mean-square values of the coefficients b_{lm} in these simulations. Note that the large-scale power is underestimated.

We then performed a maximum-likelihood analysis of each realization to answer the following question. Assuming that the recovered coefficients b_{lm} are accurate estimators of the corresponding true spherical harmonic coefficients a_{lm}, which values of the slope n and the normalization Q_{ps} fit the data best, in the sense of having maximum likelihood of

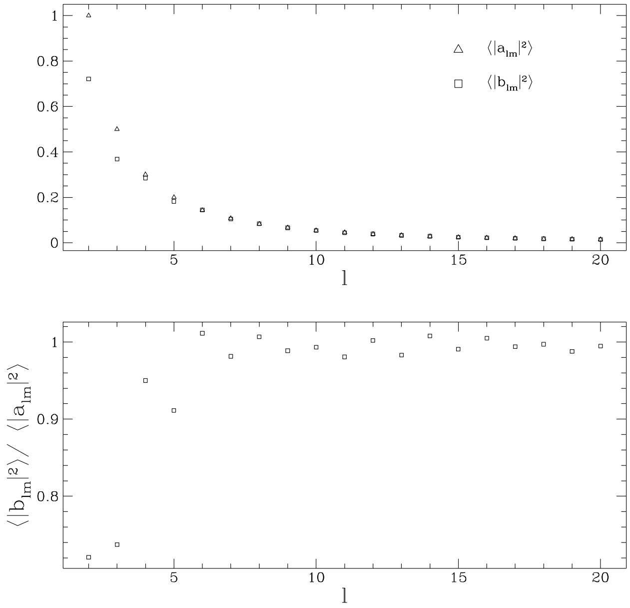

Figure 3.1: Simulated Hauser-Peebles estimates

The upper panel shows the results of Monte-Carlo simulations of the Hauser-Peebles method. $\langle |a_{lm}|^2 \rangle$ are the mean-square values of the input spherical harmonic coefficients, and $\langle |b_{lm}|^2 \rangle$ are the coefficients estimated by the Hauser-Peebles method with a region of half-width 20° about the equator removed. The lower panel shows the ratio of estimated to input coefficients. The statistical error in the estimated coefficients is approximately 1%.

generating the observed coefficients? The degree to which the answer to this question differs from the true values of n and Q_{ps} is a measure of the failure of the recovered coefficients to estimate reliably the true coefficients.

In performing such an analysis it is important to take proper account of cosmic variance. As described in Section 2.2, even with complete sky coverage, determining the a_{lm}'s for some finite range of l does not allow one to recover n and C_2 exactly. The reason is that each a_{lm} is a random variable, and the local values of these variables do not allow one to determine the ensemble-average quantities C_l. This problem is most acute for low l values: since the number of independent coefficients a_{lm} grows with increasing l, the average over all m of a_{lm} does accurately reflect the ensemble average C_l for large l. In order to separate the effect of masking from this effect, we performed simulations both with and

l_{min}	l_{max}	Cut at 20°	n	$Q^2_{\rm ps}$
2	20	Yes	1.10 ± 0.18	0.88 ± 0.30
2	20	No	0.99 ± 0.15	1.02 ± 0.27
4	20	Yes	1.02 ± 0.20	1.02 ± 0.39
4	20	No	1.01 ± 0.18	1.00 ± 0.34

Table 3.1: Monte-Carlo simulations of the Hauser-Peebles method

without an excluded region of the sky.

The results are summarized in Table 3.1. The columns labeled n and $Q^2_{\rm ps}$ show the mean values of the recovered index and normalization of the power spectrum, in units in which the input normalization is 1. When the quadrupole and octupole are excluded from the fit, masking does not significantly compromise the results. The large variances in n and $Q^2_{\rm ps}$ are almost entirely due to cosmic variance, as can be seen by comparing the results with the cut to those without the cut. Histograms showing the distribution of the estimated index n are shown in Figure 3.2.

If the quadrupole and octupole are included in the fit, the results get much worse. This is to be expected, since we have already noted that the Hauser-Peebles method fails to estimate the low-l multipoles accurately.[1]

3.3 Least-Squares Method

There is, of course, no reason to suspect that the b_{lm}'s computed by the Hauser-Peebles method will be reliable estimates of the true coefficients. Least-squares fitting seems like a more reliable method, but as we shall see it has serious difficulties.

To estimate multipole coefficients by least-squares fitting, one needs to choose a finite range of values of l. Then coefficients b_{lm} are chosen within this range so that the quantity Δ in equation (3.3) is minimized. The b_{lm}'s define a function that is "as close as possible" over the observed region to the actual function ΔT.

There is a linear relation between the actual multipole coefficients a_{lm} and the estimates b_{lm}: let $[l_{min}, l_{max}]$ be the range of l-values to be estimated. For each m between $-l_{max}$ and l_{max}, define $\tilde{W}_m$ to be the matrix whose elements are the numbers $W_{l_1 l_2 m}$, with

---

[1] It is not terribly surprising that the method fails to estimate these multipoles accurately: we never had any reason to expect otherwise.

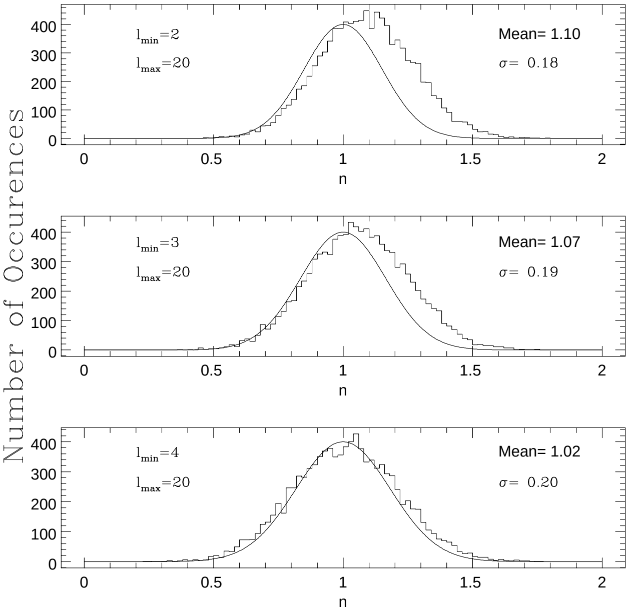

Figure 3.2: Maximum-likelihood Hauser-Peebles analysis

The results are shown of a maximum-likelihood analysis of the coefficients b_{lm} as determined by the Hauser-Peebles method. The histograms show the distribution of values of n obtained from the analysis, while the smooth curves show what one would expect from cosmic variance alone.

l_1 and l_2 ranging from $|m|$ to l_{max}. Then b_{lm} is given by

$$b_{lm} = \sum_{l_1=l_{min}}^{l_{max}} \sum_{l_2=|m|}^{\infty} (\tilde{W}_m^{-1})_{ll_1} W_{l_1 l_2 m} a_{l_2 m}. \tag{3.17}$$

(Of course, if the observed monopole and dipole are subtracted from the data, then one should use $\tilde{a}_{lm}$ from equation (12) instead of a_{lm}.)

Note that if the entire sky were observed, $\tilde{W}_m$ would be the identity matrix, and b_{lm} and a_{lm} would be identical. However, when part of the sky is masked, the b_{lm}'s become extremely bad estimates of the a_{lm}'s. In fact, each b_{lm} takes on values that are up to 10 times larger than the corresponding a_{lm}.

Figure 3.3 illustrates the problem with this method. The top panel depicts a particular realization of a Gaussian random field drawn from an $n = 1$ power spectrum. The realization includes contributions from modes with $2 \leq l \leq 20$, but the map in the figure has been smoothed by removing all contributions from modes with $l > 10$. The best-fitting coefficients b_{lm} were computed for $2 \leq l \leq 10$, and the function $\Delta T' = \sum b_{lm} Y_{lm}$ is shown in the bottom panel. As before, the observed region was taken to include all points more than 20° from the equator.

The two maps agree well over the observed region. This is to be expected: the coefficients b_{lm} are designed to do just that. However, the reconstructed map has much larger structures than it should in the masked region. This means that the total amount of power in the map is too large, and $\langle |b_{lm}|^2 \rangle$ is much larger than C_l. Figure 3.4 shows $\langle |b_{lm}|^2 \rangle$ as a function of l, assuming an $n = 1$ input power spectrum. Notice that the $\langle |b_{lm}|^2 \rangle$'s are about 100 times larger than they should be. Furthermore, the masking affects odd and even multipoles in very different ways. This behavior results from the fact that least-squares fitting is an unstable process: small changes in the input data can lead to large variations in the fit. There are regularization methods one can employ to overcome this difficulty; one such method is used in analyzing the *Relikt* CMB anisotropy data [100].

Whenever one tries to use this method to fit many multipole moments simultaneously, the results are contaminated by this effect. But what if one tries to determine just one multipole moment by this method? Figure 3.5 shows the results of Monte Carlo simulations performed to test this approach. For each of 5000 realizations, coefficients a_{lm} for $2 \leq l \leq 30$ were chosen from Gaussian distributions with widths appropriate for a Harrison-Zel'dovich power spectrum. Then the five coefficients b_{2m} were computed by least-squares fitting. The input and output quadrupole moments $Q_{in}^2 = \sum |a_{2m}|^2$ and $Q_{out}^2 = \sum |b_{2m}|^2$ are plotted in Figure 3.5. On average, the output quadrupole exceeds the input quadrupole: $Q_{out}^2 - Q_{in}^2$ has a mean value of 0.13 and a standard deviation of 0.47. This corresponds to a 1-σ uncertainty of 21% in the value of Q. Note that this uncertainty arises even when the data are free of noise. For a discussion of the effects of noise bias on such an analysis, see [75]. Other relevant considerations are presented in [167].

3.4 Correlation Function Method

This is the method that Smoot *et al.* [163] used in determining the index n and normalization $Q_{\rm ps}$ of the power spectrum. The idea is to compute the quantity

$$C(\theta) \equiv \langle \Delta T(\hat{\mathbf{r}}_1) \Delta T(\hat{\mathbf{r}}_2) \rangle_{\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2 = \cos\theta} , \tag{3.18}$$

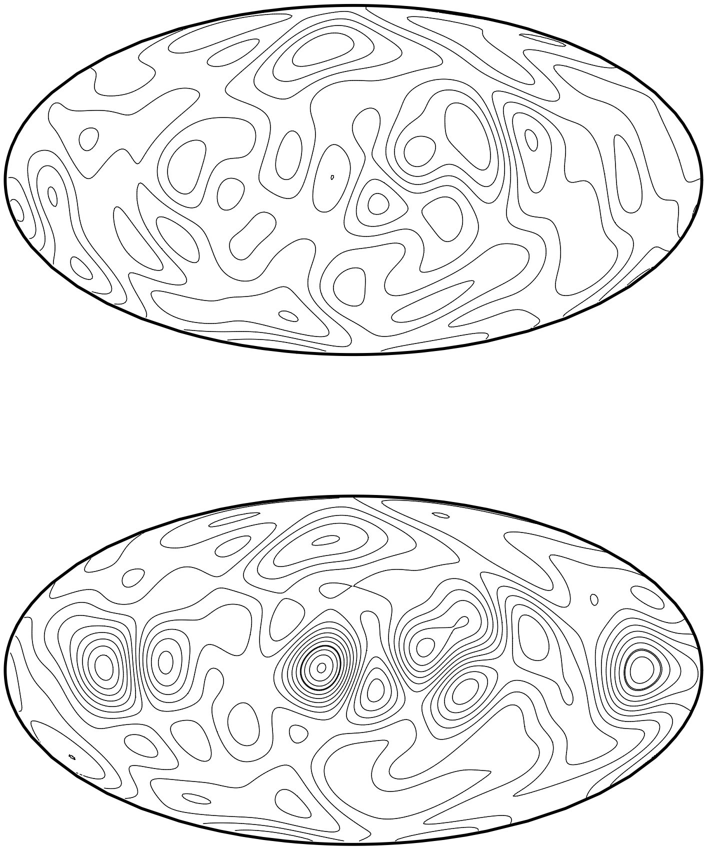

Figure 3.3: Least-squares reconstruction

The upper panel is a particular realization of a sky map, smoothed by removing all multipoles with $l > 10$. The lower panel shows the result of trying to estimate the spherical harmonic coefficients $2 \leq l \leq 10$ by least-squares fitting to this map, with the region within 20° of the equator excluded. Note the large spurious features in the masked region.

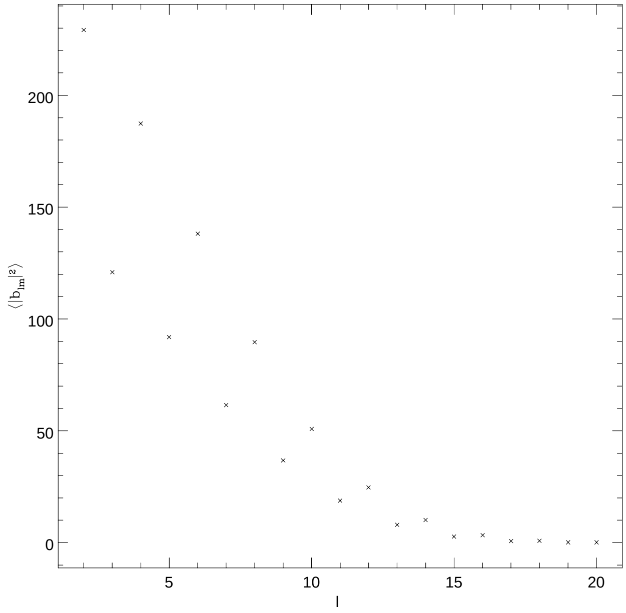

Figure 3.4: Amplitudes of least-squares coefficients

The results of trying to estimate multipoles $2-20$ by least-squares fitting excluding a 20° region are shown. The input coefficients followed a Harrison-Zel'dovich power spectrum, normalized so that the input quadrupole was 1. The estimated coefficients are too large.

and determine which values of n and $Q_{\rm ps}$ best fit it. If the average is taken over the whole sky, then $C(\theta)$ is related to a_{lm} according to equation (3.4). The remaining question is whether this relationship is significantly corrupted by incomplete sky coverage.

We performed simulations in the following way to answer this question. We made 1100 realizations of sky maps with a Harrison-Zel'dovich power spectrum, including multipoles $2 \leq l \leq 20$. These maps were divided into 1728 pixels, and the correlation function C was computed, excluding the region within 20° of the equator. The power A_l in each multipole was determined by a least-squares fit to equation (3.4), and a maximum-likelihood analysis was performed to find the best-fit index n and normalization $Q_{\rm ps}$ in the same manner as for the Hauser-Peebles method. Multipoles 19 and 20 were excluded from the fits, since they were too close to the cutoff.

Of course, it is important to perform simulations without masking in order to assess

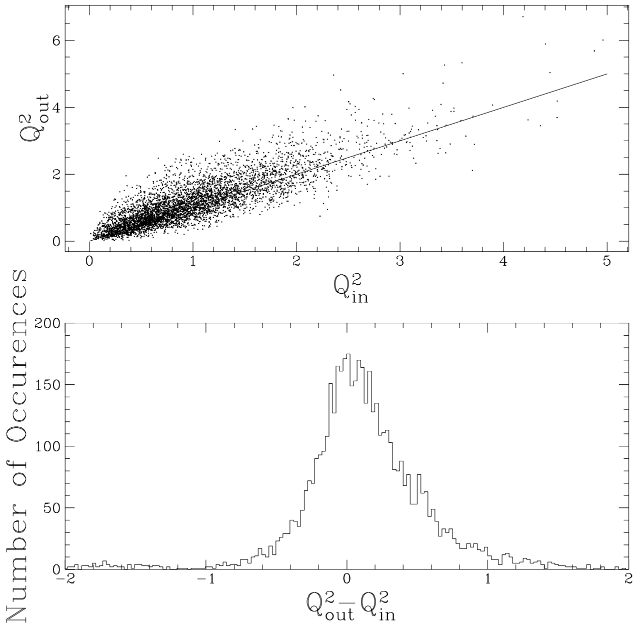

Figure 3.5: Least-squares estimates of the quadrupole

The results of trying to estimate only the quadrupole coefficients a_{2m} by least-squares fitting are shown. The solid line is the line $Q_{out} = Q_{in}$.

the effects of cosmic variance. In addition, it is of interest to introduce noise into the simulations, since the actual experimental data are quite noisy. We added Gaussian noise to each pixel with a variance of $0.29 Q^2_{\rm ps}$. This amplitude was chosen in order to make the statistical error in $C(\theta)$ approximately the same magnitude as the error bars in the Smoot *et al.* [163] correlation functions. (It is important to note, however, that our procedure does not mimic the analysis performed on the actual COBE data sufficiently precisely to expect the results to match in detail for the noisy data sets.)

The results of our analysis are shown in Table 3.2. It appears from these results that introducing a 20° cut increases the uncertainty in n and $Q_{\rm ps}$, but does not bias the results in either direction. It may be of interest to use the correlation function to determine the power in individual multipoles, rather than to attempt to constrain the power spectrum. The results of such an exercise are shown in Figure 3.6. Here the quadrupole moment determined from the correlation function is compared with the input quadrupole in realizations that

Cut at 20°	Noise	n	$Q^2_{\rm ps}$
No	No	1.01 ± 0.26	1.04 ± 0.39
Yes	No	0.96 ± 0.36	1.16 ± 0.54
No	Yes	0.97 ± 0.73	1.10 ± 0.61
Yes	Yes	1.10 ± 0.96	1.10 ± 0.78

Table 3.2: Monte-Carlo simulations of the correlation function method

included a cut at 20°, but did not include noise. $Q^2_{out} - Q^2_{in}$ has a mean value of 0.03, and a standard deviation of 0.45. This corresponds to a 20% uncertainty in Q_{out}.

3.5 Conclusions

There are two types of questions that one might consider asking when presented with large-angular-scale CMB anisotropy data. The first type is concerned with characterizing the anisotropy as seen from our particular vantage point. Such questions include attempts to determine the values of the coefficients a_{lm}; as we have seen, some such questions are doomed to failure by incomplete sky coverage. Indeed, the uncertainty in estimating any particular coefficient a_{lm} is formally infinite, unless assumptions are made that place bounds on the other coefficients [167]. If we assume that the input power spectrum is of roughly the Harrison-Zel'dovich form, we have seen that either least-squares fitting or computation of the correlation function can produce estimates of the total quadrupole power. However, these methods are valid only if we assume that the power spectrum generating the fluctuations has a particular form, and even then there are extremely large residual uncertainties. Masking induces an uncertainty of about 20% (1-σ) in estimating the quadrupole, and this uncertainty will invariably be augmented by noise in the data. It is also important to note that least-squares fitting does tend to bias the estimated quadrupole upwards.

It is clear from these results that one should be very skeptical of attempts to use the COBE estimate of the local quadrupole Q to normalize the spectrum of fluctuations in cosmological models. Q cannot be determined accurately without complete sky coverage, and so it makes a poor choice of normalization. In addition, even if Q were accurately determined, it would still have a very high intrinsic cosmic variance [2, 187].

The second type of question involves using our local measurements of CMB anisotropy to place constraints on cosmological models. Attempts to determine n and $Q_{\rm ps}$ fall

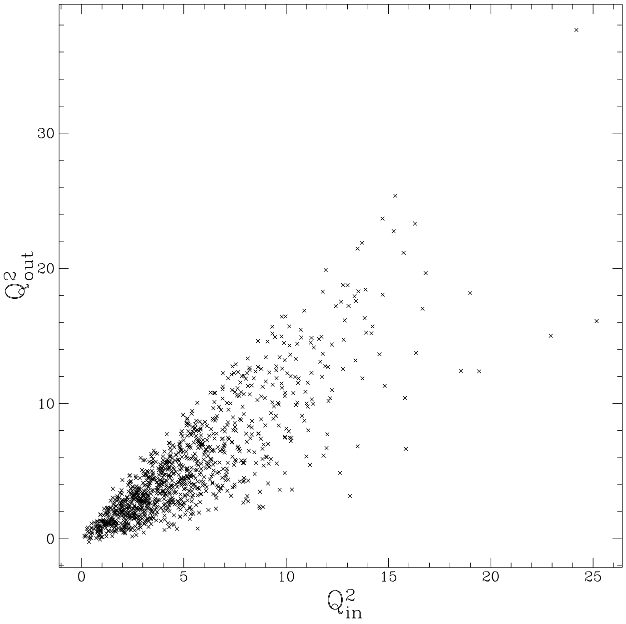

Figure 3.6: Estimating the quadrupole from the correlation function

The quadrupole moment obtained from the masked correlation function is plotted against the input quadrupole moment.

into this category. The Hauser-Peebles method seems to be a reliable tool for this task (although one must take care to correct for the fact that the method does not estimate the low multipoles in an unbiased way). Unlike the correlation function method, it suffers essentially no degradation from incomplete sky coverage. As the noise in the COBE data continues to decrease, the combined effects of cosmic variance and incomplete sky coverage will place limits on the accuracy with which these parameters can be determined: n will necessarily have a 1-σ uncertainty of order 0.2, and $Q_{\rm ps}$ will have an uncertainty of order 20%. In order to reduce these uncertainties, it will be necessary to combine large-scale experiments like COBE with experiments that probe smaller angular scales. (*e.g.*, [64, 51].) These experiments have their own limitations caused by cosmic variance and incomplete sky coverage, but the uncertainties can be significantly reduced by increasing the fraction of sky covered by the experiments [154].

Chapter 4

The Karhunen-Loève Transform

> Transform me then, and to your power I'll yield.
> William Shakespeare

4.1 Likelihood Analysis

In this chapter and the next we will attempt to use the DMR data to test hypotheses about the power spectrum C_l. We will also attempt the related task of parameter estimation: we will assume that the power spectrum is a member of a few-parameter family, and we will try to determine the best-fitting values of those parameters. We will adopt an approach that is Bayesian in spirit, although we will comment on the merits of a complementary frequentist approach. A useful textbook on these subjects is that of Berger [15].

In Chapter 2, we adopted the hypothesis that the CMB anisotropy forms an isotropic Gaussian random field. If we also assume that the noise is Gaussian, then we know the probability distribution function of the data $\vec{d}$ given a theoretical power spectrum $\mathbf{C}$. Following the notation of Chapter 2, it is

$$f(\vec{d}\,|\,\mathbf{C}) = \frac{1}{(2\pi)^{\mathcal{N}/2}\det^{1/2}\mathcal{M}}\exp\left(-\tfrac{1}{2}\vec{d}^T\cdot\mathcal{M}^{-1}\cdot\vec{d}\right), \tag{4.1}$$

where the data covariance matrix $\mathcal{M}$ is

$$\mathcal{M} = \langle\vec{d}\vec{d}^T\rangle = \mathbf{Y}\cdot\bar{\mathbf{C}}\cdot\mathbf{Y}^T + \mathbf{N}. \tag{4.2}$$

Here $\bar{\mathbf{C}} = \mathbf{W}\cdot\mathbf{C}\cdot\mathbf{W}$, $\mathbf{W}$ is the window function, and $\mathbf{N}$ is the noise covariance matrix. We will regard f as a function of $\mathbf{C}$ rather than as a function of the data $\vec{d}$: after all, $\vec{d}$ is fixed and $\mathbf{C}$ is unknown. We will call it the *likelihood* $L(\mathbf{C})$. When we wish to estimate some set of parameters $\vec{q}$ (which may of course be a set containing only one element), we will often choose as an estimate the value of $\vec{q}$ for which $L(\mathbf{C}(\vec{q}))$ is maximized.[1]

When we wish to establish credible regions around a parameter estimate, we will do so in the following way. Bayes's theorem states that the probability distribution of a set of parameters $\vec{q}$ given the data $\vec{d}$ is

$$f(\vec{q}\,|\,\vec{d}) \propto f(\vec{d}\,|\,\vec{q})f(\vec{q}). \tag{4.3}$$

The first term is simply the likelihood $L(\vec{q})$, and the second is some prior distribution. We will generally adopt uniform prior distributions, taking $f(\vec{q})$ to be a constant. Then the likelihood $L(\vec{q})$ is identical to the probability distribution of the parameters $\vec{q}$ (up to an overall normalization). A *Bayesian credible region* in $\vec{q}$-space is then simply a volume V bounded by a surface of constant L such that

$$\int_V L(\vec{q})\,d\vec{q} = c\int L(\vec{q})\,d\vec{q}, \tag{4.4}$$

where c is the confidence level associated with the region, and the integral on the right-hand side extends over all of $\vec{q}$ space.[2]

These prescriptions tell us how to perform our parameter estimates. Of course, we have no guarantee that these estimates will be unbiased. (In the case of Gaussian statistics, parameter estimates are asymptotically unbiased; however, since we deal with finite sample sizes, this guarantee does not apply.) If we are concerned about bias, we therefore have no recourse but to perform Monte Carlo simulations to test for it.

Before we can do any of this we need to account properly for the fact that we do not have any useful information about the monopole and dipole anisotropy. The natural way to do this in a Bayesian context is to marginalize over the unknown modes as follows. Let $\mathbf{Z}$ be a matrix containing the first four columns of $\mathbf{Y}$, representing the monopole and dipole.[3] Then our data contains not only the cosmic signal $\mathbf{Y}\cdot\vec{a}$ and noise $\vec{n}$, but also some unknown contribution $\mathbf{Z}\cdot\vec{b}$ due to the monopole and dipole. Here $\vec{b}$ is a four-dimensional vector about which we know nothing.

---

[1] Often for compactness of notation we will think of L as a function of the parameters $\vec{q}$ in our family of power spectra and denote $L(\mathbf{C}(\vec{q}))$ by $L(\vec{q})$.

[2] Note that this procedure is completely general, and in no way depends upon the assumption of Gaussian statistics. Of course, if we do not make the assumption of Gaussianity, then equation (4.1) is not valid, and we need to replace it with something else.

[3] If you choose to throw away the quadrupole as well, ignoring the advice in Section 2.7, then $\mathbf{Z}$ should contain nine columns.

We did perform a least-squares fit to the monopole and dipole in the data, but in the absence of complete sky coverage we have no guarantee that we removed the correct monopole and dipole. In fact, we should be concerned that we covertly removed contributions from other multipoles at the same time. The Gaussian likelihood formula (4.1) really applies to the data that we would get if we had correctly removed the true monopole and dipole. That is, equation (4.1) should really read

$$f(\vec{d}\,|\,\mathbf{C},\vec{b}) = \frac{1}{(2\pi)^{\mathcal{N}/2}\det^{1/2}\mathcal{M}}\exp\left(-\tfrac{1}{2}(\vec{d}-\mathbf{Z}\cdot\vec{b})^T\cdot\mathcal{M}^{-1}\cdot(\vec{d}-\mathbf{Z}\cdot\vec{b})\right). \tag{4.5}$$

Bayes's theorem tells us that the actual likelihood is simply

$$f(\vec{d}\,|\,\mathbf{C}) \propto \int d^4b\, f(\vec{d}\,|\,\mathbf{C},\vec{b})f(\vec{b}), \tag{4.6}$$

where $f(\vec{b})$ is the prior distribution of $\vec{b}$. Since we have no information about $\vec{b}$, we take this distribution to be an uninformative constant. The marginalized likelihood function is therefore

$$L(\mathbf{C}) = \frac{1}{(2\pi)^{\mathcal{N}/2}\det^{1/2}\mathcal{M}}\int d^4b\,\exp\left(-\tfrac{1}{2}(\vec{d}-\mathbf{Z}\vec{b})^T\cdot\mathcal{M}^{-1}\cdot(\vec{d}-\mathbf{Z}\vec{b})\right). \tag{4.7}$$

We can evaluate this integral by completing the square in $\vec{b}$ to get

$$L(\mathbf{C}) = \frac{1}{(2\pi)^{(\mathcal{N}-4)/2}\det^{1/2}\mathcal{M}\,\det^{1/2}(\mathbf{Z}^T\cdot\mathcal{M}^{-1}\cdot\mathbf{Z})^{-1}}\exp\left(-\tfrac{1}{2}((\mathbf{P}\cdot\vec{d})^T\cdot\mathcal{M}^{-1}\cdot\mathbf{P}\cdot\vec{d})\right), \tag{4.8}$$

where the projection operator $\mathbf{P} = \mathbf{1}-\mathbf{Z}\cdot(\mathbf{Z}^T\cdot\mathcal{M}^{-1}\cdot\mathbf{Z})^{-1}\cdot\mathbf{Z}^T\cdot\mathcal{M}^{-1}$. This operator orthogonally projects out the monopole and dipole contributions to $\vec{d}$, where orthogonality is defined with respect to the inner product $\langle\vec{x},\vec{y}\rangle = \vec{x}^T\cdot\mathcal{M}^{-1}\cdot\vec{y}$.

Of course, all of these considerations are relevant only if we can compute the likelihood function. The likelihood $L(\mathbf{C})$ is rather awkward to compute, as it involves inverting the $\mathcal{N}\times\mathcal{N}$ matrix $\mathcal{M}$. This task has been performed for the family of models based on the Sachs-Wolfe power spectrum (2.29) [174], but these computations are prohibitively slow for exploring a large parameter space, especially if we want to perform Monte Carlo simulations to test for bias. In the next section, therefore, we will propose a way of approximating the likelihood L by projecting the data vector $\vec{d}$ onto a subspace of smaller dimension. Our goal is to choose this subspace in such a way that the likelihood function based on the new "compressed" data vector contains almost all of the useful information in the original likelihood function.

The idea behind this sort of data compression is quite simple. If the pixels in our map covered the sky uniformly, we could estimate the coefficients a_{lm} directly:

$$\hat{a}_{lm} = \frac{4\pi}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} d_i Y_{lm}(\hat{\mathbf{r}}_i), \tag{4.9}$$

where d_i is the ith data point, and $\hat{\mathbf{r}}_i$ is the location of the ith pixel on the sky.[4] We could then compute likelihoods using the estimates $\hat{a}_{lm}$ instead of $\vec{d}$. Because the COBE data are insensitive to small angular scales, we could consider only modes below some $l_{\max}$, making the process of computing likelihoods more efficient than computing them from the raw data.

As we saw in the previous chapter, since our data set does not cover the whole sky, this procedure does not yield good estimates of the coefficients a_{lm}. We can still compute the sum in equation (4.9), but the results no longer have a direct physical interpretation. We know the probability distribution of the $\hat{a}_{lm}$'s (like most other probability distributions in this dissertation, it's a multivariate Gaussian), and so we can still compute likelihoods; however, since $\hat{a}_{lm}$ is not related to a_{lm} in any particular way, there no longer seems to be any compelling reason to use this procedure.

However, the procedure of expanding the data set in terms of a suitable set of basis functions is still fruitful. Let us choose a set of basis functions $\{\alpha_1, \ldots, \alpha_K\}$ from the sky to the real numbers. We will define a K-dimensional vector $\vec{y}$ by

$$y_a = \sum_{i=1}^{\mathcal{N}} d_i \alpha_a(\hat{\mathbf{r}}_i), \tag{4.10}$$

and use $\vec{y}$ instead of $\vec{d}$ to compute likelihoods. (In effect, we are simply replacing the spherical harmonics in equation (4.9) by other functions.) Specifically, we will compute

$$L_y = \frac{1}{(2\pi)^{K/2} \det^{1/2} \mathbf{M}} \exp\left(-\tfrac{1}{2}\vec{y}^T \cdot \mathbf{M}^{-1} \cdot \vec{y}\right), \tag{4.11}$$

where $\mathbf{M} = \langle \vec{y} \cdot \vec{y}^T \rangle$ is the covariance matrix of $\vec{y}$.

If we choose the functions α appropriately, then L_y may prove to be a good estimate of the likelihood even if K is considerably less than $\mathcal{N}$. In the next section we will propose an optimal method based on the Karhunen-Loève transformation [99, 176] for choosing the functions α_a in order to achieve this goal.

---

[4] These estimates are actually of the beam-smoothed coefficients $W_{lm} a_{lm}$.

4.2 An Optimization Problem

Let us remind the reader of the notation established in Section 2.6. The data vector $\vec{d}$ contains both signal and noise contributions:

$$\vec{d} = \mathbf{Y} \cdot \vec{a} + \vec{n}. \tag{4.12}$$

In the cosmological models we are considering, the quantities a_μ are independent Gaussian random variables with zero mean. To a very good approximation, so are the quantities n_i. The covariance matrices of these quantities are

$$\langle a_\mu a_\nu \rangle = C_l \delta_{\mu\nu} \equiv C_{\mu\nu}, \tag{4.13}$$

$$\langle n_i n_j \rangle = \sigma_i^2 \delta_{ij} \equiv N_{ij}, \tag{4.14}$$

$$\langle a_\mu n_i \rangle = 0. \tag{4.15}$$

Throughout this section, we will ignore beam-smoothing by setting the window function $\mathbf{W} = \mathbf{1}$. This is purely to simplify the notation and makes no essential difference to the problem. If you absolutely must know where the $\mathbf{W}$'s go, simply replace $\vec{a}$ by $\mathbf{W} \cdot \vec{a}$ and $\mathbf{C}$ by $\mathbf{W} \cdot \mathbf{C} \cdot \mathbf{W}$ everywhere in this section. We also ignore the small off-diagonal elements of the noise covariance matrix $\mathbf{N}$ when performing the Karhunen-Loève transformation. If we like, we can put them back in before computing likelihoods, although it turns out to make virtually no difference whether we do or not.

Define a matrix $\mathbf{A}$ by $A_{ai} = \alpha_a(\hat{\mathbf{r}}_i)$. Then the compressed data vector is $\vec{y} = \mathbf{A} \cdot \vec{d}$, and its covariance matrix is

$$\mathbf{M} \equiv \langle \vec{y} \cdot \vec{y}^T \rangle = \langle \mathbf{A} \cdot \vec{d} \cdot \vec{d}^T \cdot \mathbf{A}^T \rangle = \mathbf{A} \cdot \mathcal{M} \cdot \mathbf{A}^T, \tag{4.16}$$

and the likelihood of $\vec{y}$ for a given power spectrum is given by

$$-2 \ln L = \ln \det \mathbf{M} + \vec{y}^T \cdot \mathbf{M}^{-1} \cdot \vec{y} = \mathrm{Tr}\left(\mathbf{M}^{-1} \cdot \vec{y} \cdot \vec{y}^T + \ln \mathbf{M} \right). \tag{4.17}$$

We need to choose our basis functions α_a, or equivalently the matrix $\mathbf{A}$. We will adopt the following criterion. Suppose that the true angular power spectrum is $\mathbf{C}_0$, and suppose that $\mathbf{C}(q)$ is a one-parameter family of power spectra, with $\mathbf{C}(0) = \mathbf{C}_0$. We shall see below that the ensemble-average likelihood function $\langle \ln L \rangle$, considered as a function of q, has a maximum at $q = 0$. We will choose our basis functions to satisfy the criterion that $\langle \ln L \rangle$ should decrease as fast as possible as we vary q away from its maximum at $q = 0$.

In other words, we will choose $\mathbf{A}$ to maximize our expected rejection power for incorrect values of q.

If we Taylor expand $\langle\ln L\rangle$ about $q=0$, we see that this is equivalent to minimizing $\gamma \equiv d^2\langle\ln L\rangle/dq^2$. From equation (4.17) we have

$$-2\ln L = \mathrm{Tr}\left(\mathbf{A}^T\cdot(\mathbf{A}\cdot\mathcal{M}\cdot\mathbf{A}^T)^{-1}\cdot\mathbf{A}\cdot\mathcal{M}_0 + \ln\mathbf{A}\cdot\mathcal{M}\cdot\mathbf{A}^T\right), \tag{4.18}$$

where $\mathcal{M}_0$ is the covariance matrix $\mathcal{M}$ evaluated at the true value $q=0$ and we have made use of the fact that $\mathcal{M}_0 = \langle\vec{d}\cdot\vec{d}^T\rangle$. Denoting derivatives with respect to q by primes, we have

$$\begin{aligned}\langle(\ln L)'\rangle = \tfrac{1}{2}\mathrm{Tr}\Big(&\mathbf{A}\cdot\mathcal{M}_0\cdot\mathbf{A}^T\cdot(\mathbf{A}\cdot\mathcal{M}\cdot\mathbf{A}^T)^{-1}\cdot\mathbf{A}\cdot\mathcal{M}'\cdot\mathbf{A}^T\cdot(\mathbf{A}\cdot\mathcal{M}\cdot\mathbf{A}^T)^{-1}\\ &-(\mathbf{A}\cdot\mathcal{M}\cdot\mathbf{A}^T)^{-1}\cdot\mathbf{A}\cdot\mathcal{M}'\cdot\mathbf{A}^T\Big).\end{aligned} \tag{4.19}$$

This derivative vanishes when we set $q=0$ (and hence $\mathcal{M}=\mathcal{M}_0$), showing that the average value of $\ln L$ is maximal at the correct point $q=0$. Differentiating (4.19) again and setting $q=0$, we find that

$$\gamma = -\mathrm{Tr}\left[((\mathbf{A}\cdot\mathcal{M}_0\cdot\mathbf{A}^T)\cdot(\mathbf{A}\cdot\mathcal{M}_0'\cdot\mathbf{A}^T))^2\right]. \tag{4.20}$$

To find $\mathbf{A}$, we set $\partial\gamma/\partial A_{ij}=0$. After some unenlightening algebra, we get

$$\mathbf{A}\cdot\mathcal{M}_0' = (\mathbf{A}\cdot\mathcal{M}_0'\cdot\mathbf{A}^T)\cdot(\mathbf{A}\cdot\mathcal{M}_0\cdot\mathbf{A}^T)^{-1}\cdot\mathbf{A}\cdot\mathcal{M}_0. \tag{4.21}$$

Although it might not look like it, this is in fact a generalized eigenvalue equation for the rows of $\mathbf{A}$. In order to see this, use Cholesky decomposition to write $\mathcal{M}_0 = \mathbf{L}\cdot\mathbf{L}^T$, and set $\hat{\mathbf{A}} = \mathbf{A}\cdot\mathbf{L}$ and $\hat{\mathcal{M}}' = \mathbf{L}^{-1}\cdot\mathcal{M}_0'\cdot(\mathbf{L}^T)^{-1}$. Then equation (4.21) becomes

$$\hat{\mathbf{A}}\cdot\hat{\mathcal{M}}' = \hat{\mathbf{A}}\cdot\hat{\mathcal{M}}'\cdot\hat{\mathbf{A}}^T\cdot(\hat{\mathbf{A}}\cdot\hat{\mathbf{A}}^T)^{-1}\cdot\hat{\mathbf{A}} = \hat{\mathbf{A}}\cdot\hat{\mathcal{M}}'\cdot\mathbf{P}. \tag{4.22}$$

The matrix $\mathbf{P}$ is the projection operator that orthogonally projects vectors onto the K-dimensional row space of $\hat{\mathbf{A}}$. This equation states that the matrix $\hat{\mathbf{A}}\cdot\hat{\mathcal{M}}'$ is unchanged under this projection, or, equivalently, that the row space of $\hat{\mathbf{A}}\cdot\hat{\mathcal{M}}'$ is the same as that of $\hat{\mathbf{A}}$. That means that

$$\hat{\mathbf{A}}\cdot\hat{\mathcal{M}}' = \mathbf{Q}\cdot\hat{\mathbf{A}} \tag{4.23}$$

for some $K\times K$ matrix $\mathbf{Q}$. Since $\hat{\mathcal{M}}'$ is symmetric, we can write it in the form $\mathbf{R}\cdot\boldsymbol{\Lambda}\cdot\mathbf{R}^T$ where the columns of the orthogonal matrix $\mathbf{R}$ are the eigenvectors of $\hat{\mathcal{M}}'$, and the diagonal matrix $\mathit{\Lambda}$ contains the corresponding eigenvalues. Let us assume for simplicity that the eigenvalues are all distinct, although the final result does not depend on this.[5] Then if we set $\mathbf{B} = \hat{\mathbf{A}} \cdot \mathbf{R}$, we find that

$$\mathbf{B} \cdot \mathit{\Lambda} = \mathbf{Q} \cdot \mathbf{B}. \tag{4.24}$$

If $\vec{b}_i$ is the ith column of $\mathbf{B}$, then we have $\lambda_i \vec{b}_i = \mathbf{Q} \cdot \vec{b}_i$. That is, $\vec{b}_i$ is either an eigenvector of $\mathbf{Q}$ or it is zero. Since $\mathbf{Q}$ can have at most K distinct eigenvalues, at least $\mathcal{N} - K$ columns of $\mathbf{B}$ must be zero. Let us rearrange the order of the eigenvectors in $\mathbf{R}$ so that these are the last $\mathcal{N} - K$ columns of $\mathbf{B}$. Then

$$\hat{\mathbf{A}} = \mathbf{B} \cdot \mathbf{R}^T = \mathbf{B}_* \cdot \mathbf{R}_*^T, \tag{4.25}$$

where $\mathbf{B}_*$ is the $K \times K$ matrix containing the first K columns of $\mathbf{B}$ (that is, $\mathbf{B}_*$ is $\mathbf{B}$ shorn of its zeroes), and $\mathbf{R}_*$ is the $\mathcal{N} \times K$ matrix containing the first K columns of $\mathbf{R}$.

Recall that our goal was to find a matrix $\mathbf{A}$ satisfying equation (4.21). It is easy to check that if $\mathbf{A}$ satisfies this equation, then $\mathbf{Z} \cdot \mathbf{A}$ also satisfies it for any nonsingular matrix $\mathbf{Z}$.[6] So, from equation (4.25), we can conclude that our solution to equation (4.21) can be written in the form $\hat{\mathbf{A}} = \mathbf{R}_*^T$. That is, the rows of $\hat{\mathbf{A}}$ are eigenvectors of $\hat{\mathcal{M}}'$.[7] Converting from hatted to hatless matrices, we have our final result: The rows $\vec{\alpha}_a$ of $\mathbf{A}$ are solutions to the generalized eigenvalue equation

$$\mathcal{M}_0' \cdot \vec{\alpha}_a = \lambda_a \mathcal{M}_0 \cdot \vec{\alpha}_a. \tag{4.26}$$

Furthermore,

$$\gamma = -\sum_{a=1}^{K} \lambda_a^2, \tag{4.27}$$

which means that we should choose the rows of $\mathbf{A}$ to be the solutions of equation (4.26) with the largest eigenvalues (in absolute value).

---

[5] If some eigenvalues are degenerate, then we have additional freedom in choosing the columns of $\mathbf{R}$. By employing this additional freedom judiciously, we can make the argument we are about to give work even in this case. The details can be worked out easily by the reader.

[6] In fact, the entire likelihood function (4.18) is unchanged up to an overall multiplicative constant if $\mathbf{A}$ is replaced by $\mathbf{Z} \cdot \mathbf{A}$. One way to say this is that the likelihood cares only about the subspace spanned by the rows of $\mathbf{A}$, not about the basis with which we choose to specify that subspace.

[7] We have implicitly assumed here that $\mathbf{B}$ is nonsingular. This is the same as restricting our attention to matrices $\mathbf{A}$ that have maximal rank K. This seems like a perfectly reasonable thing to do: it would be hard to believe that a matrix of some rank less than K could do better than any matrix of rank K. In fact, it is easy to prove that this can never happen. Given a matrix of rank less than K, it is always possible to "extend" it to rank K in a way that does not decrease γ. This is proved in [174].

We now have a definite procedure for choosing our matrix $\mathbf{A}$. We solve equation (4.26) and order the solutions from largest to smallest eigenvalues. These are the rows of $\mathbf{A}$. Furthermore, we have a criterion for choosing a value for K, the number of basis functions we need to use: once the eigenvalues λ_a have gotten very small, including more modes does not significantly increase the rejection power γ.

Let us try to get an intuitive understanding of these basis functions. Suppose that the parameter q is the overall normalization of the power spectrum, so that $\mathbf{C}(q) = (1+q)\mathbf{C}_0$. Then if we separate the covariance matrix into signal and noise contributions,

$$\mathcal{M}_0 = \mathcal{M}_{\rm sig} + \mathcal{M}_{\rm noise} = \mathbf{Y}\cdot\mathbf{C}_0\cdot\mathbf{Y}^T + \mathbf{N}, \tag{4.28}$$

then $\mathcal{M}_0' = \mathcal{M}_{\rm sig}$, and equation (4.29) becomes

$$\mathcal{M}_{\rm noise}^{-1}\cdot\mathcal{M}_{\rm sig}\cdot\vec{\alpha}_a = \hat{\lambda}_a\vec{\alpha}_a, \tag{4.29}$$

where $\hat{\lambda}_a = \lambda_a/(1-\lambda_a)$. $\vec{\alpha}$ is therefore an eigenmode of $\mathcal{M}_{\rm noise}^{-1}\cdot\mathcal{M}_{\rm sig}$, or to use Bond's terminology [20, 21], it is "an eigenmode of the signal-to-noise ratio." Each vector $\vec{\alpha}$ corresponds to a direction in the $\mathcal{N}$-dimensional vector space in which $\vec{d}$ lives. When we compute the reduced data vector $\vec{y}$, we are projecting $\vec{d}$ along each of these directions. Equation (4.29) simply says that we should choose directions in that space that have large contributions from the signal and small contributions from the noise.

Before we can compute the matrix $\mathbf{A}$, we need to choose a fiducial power spectrum $\mathbf{C}_0$. We would like to choose a power spectrum that is close to the true power spectrum. Of course, we don't know the true power spectrum: if we did, there would be no need for this analysis. However, we shall see below that our results are not very sensitive to the particular choice we make for $\mathbf{C}_0$, as long as it falls off fairly rapidly with l. (Remember that in this section, C_l is the beam-smoothed power spectrum $\bar{C}_l$, which means that we expect it to fall off roughly as a Gaussian in l.) We have chosen to use a pure Sachs-Wolfe $n=1$ spectrum,

$$C_l \propto \frac{1}{l(l+1)}, \tag{4.30}$$

for our fiducial power spectrum. We will discuss the effects of varying this choice below.

I can almost hear you, dear reader, muttering to yourself, "Now wait a minute. We got into this mess because we didn't want to invert the big matrix $\mathcal{M}$. Now you're telling me I have to find the eigenvectors of $\mathcal{M}$, which is *much* harder!" Things really aren't as bad as all that. First of all, our objection was only to *repeated* inversions of $\mathcal{M}$, and we only have to solve the eigenvector problem once. Second, there is a trick that enables us to find the eigenvectors we need without actually diagonalizing $\mathcal{M}$.

Let us suppose first that the parameter q with respect to which we are optimizing is the overall normalization. The matrix $\mathbf{Y}$ is in principle $\mathcal{N} \times \infty$, but in practice after some column L the contribution of subsequent columns is negligible. Let us truncate $\mathbf{Y}$ at this point, making it an $\mathcal{N} \times L$ matrix. Correspondingly, $\mathbf{C}$ will be truncated to $L \times L$. By a little bit of manipulation we can reduce our eigenvector problem from an $\mathcal{N} \times \mathcal{N}$ problem to an $L \times L$ one. (Of course, we will have to check to see whether we have introduced any significant error by truncating $\mathbf{Y}$.)

Equation (4.29) says

$$\mathbf{Y} \cdot \mathbf{C} \cdot \mathbf{Y}^T \cdot \vec{\alpha} = \hat{\lambda} \mathbf{N} \cdot \vec{\alpha}. \tag{4.31}$$

Let $\vec{\beta} = \mathbf{N}^{1/2} \cdot \vec{\alpha}$. (Any square root of $\mathbf{N}$ will do, but since $\mathbf{N}$ is diagonal, there is a natural choice.) Setting $\mathbf{Z} = \mathbf{N}^{-1/2} \cdot \mathbf{Y} \cdot \mathbf{C}^{1/2}$, equation (4.31) becomes

$$\mathbf{Z} \cdot \mathbf{Z}^T \cdot \vec{\beta} = \hat{\lambda} \vec{\beta}. \tag{4.32}$$

It is clear from this equation that $\vec{\beta}$ lies in the row-space of $\mathbf{Z}$ whenever $\hat{\lambda} \neq 0$. Thus, there is some $\vec{c}$ such that $\vec{\beta} = \mathbf{Z} \cdot \vec{c}$. If we choose $\vec{c}$ to be a solution of the eigenvalue equation

$$\mathbf{Z}^T \cdot \mathbf{Z} \cdot \vec{c} = \hat{\lambda} \vec{c}, \tag{4.33}$$

then the result will be a solution of our original eigenvalue problem. Furthermore, by simply counting solutions to the two eigenvalue equations we see that all of the solutions of equation (4.31) with nonzero eigenvalues are obtained by this process. Of course, the original eigenvalue equation (4.29) has a full set of $\mathcal{N}$ nonzero eigenvalues. By truncating $\mathbf{Y}$ we have reduced the rank of $\mathcal{M}_{\text{sig}}$ and forced some eigenvalues that should be nonzero to zero. However, we are only interested in the eigenvectors with large eigenvalues, and it is reasonable to suppose that these are relatively unaffected by truncation. We shall see below that this is in fact the case.

When q is some other parameter, such as the slope n in equation (2.29), the same substitutions yield the eigenvalue equation[8]

$$(\mathbf{1}+\mathbf{C}\cdot\mathbf{Z}^T\cdot\mathbf{Z})^{-1}\cdot\mathbf{C}'\cdot\mathbf{Z}^T\cdot\mathbf{Z}\cdot\vec{c}=\lambda\vec{c}. \tag{4.34}$$

4.3 Transforming the DMR Data

In this section we will perform the Karhunen-Loève transform on the weighted-average DMR data shown in Figure 2.2 and give some preliminary results. Let us adopt the fiducial power spectrum (4.30), and take the normalization Q_{ps}^2 to be the parameter q with respect to which we optimize. Then we can solve the eigenvalue equation (4.33) to find the optimal basis vectors $\vec{\alpha}_a$. We can of course consider these vectors as functions on the sphere evaluated at the locations of the data points.

In Figure 4.1 we show a sample of these eigenmodes. The modes with large eigenvalues (*i.e.*, the ones with high signal-to-noise ratios) tend to be slowly varying and hence probe large angular scales. This simply reflects the fact that the fiducial power spectrum (and, we suspect, the real power spectrum) decreases with increasing l. Note that some modes occupy primarily only one hemisphere. These modes are always nearly degenerate with another mode that similarly occupies the other hemisphere.

We can expand each mode in spherical harmonics with coefficients b_{lm}, and assign an "effective l" to each mode by computing the centroid of l with respect to the squares of the coefficients:

$$l_{\mathrm{eff}}=\frac{\sum_{l,m} l b_{lm}^2}{\sum_{l,m} b_{lm}^2}. \tag{4.35}$$

(The modes tend to have fairly narrow widths in l-space, so the exact definition of l_{eff} is not too important.) Figure 4.2 shows these effective l's as a function of mode number with the modes sorted in decreasing order by eigenvalue.

We have computed the eigenmodes using the method described at the end of the previous section, using two different cutoffs $l_{\max}$ (and hence two different values of the number $L=(l_{\max}+1)^2$ of columns we keep in $\mathbf{Y}$). We show both cutoffs in Figure 4.2. It is reasonable to guess that the eigenmodes are uncorrupted by the cutoff in l as long as we restrict our attention to modes that lie well before the turnover in Figure 4.2, say around

---

[8] Unlike equation (4.33), this equation is not cast in the form of a generalized eigenvalue problem involving symmetric matrices. This means that one must pull a different and somewhat slower routine from one's package of canned eigenvector routines. The present author, incidentally, is partial to the NAG library [128] for solving eigenvalue problems.

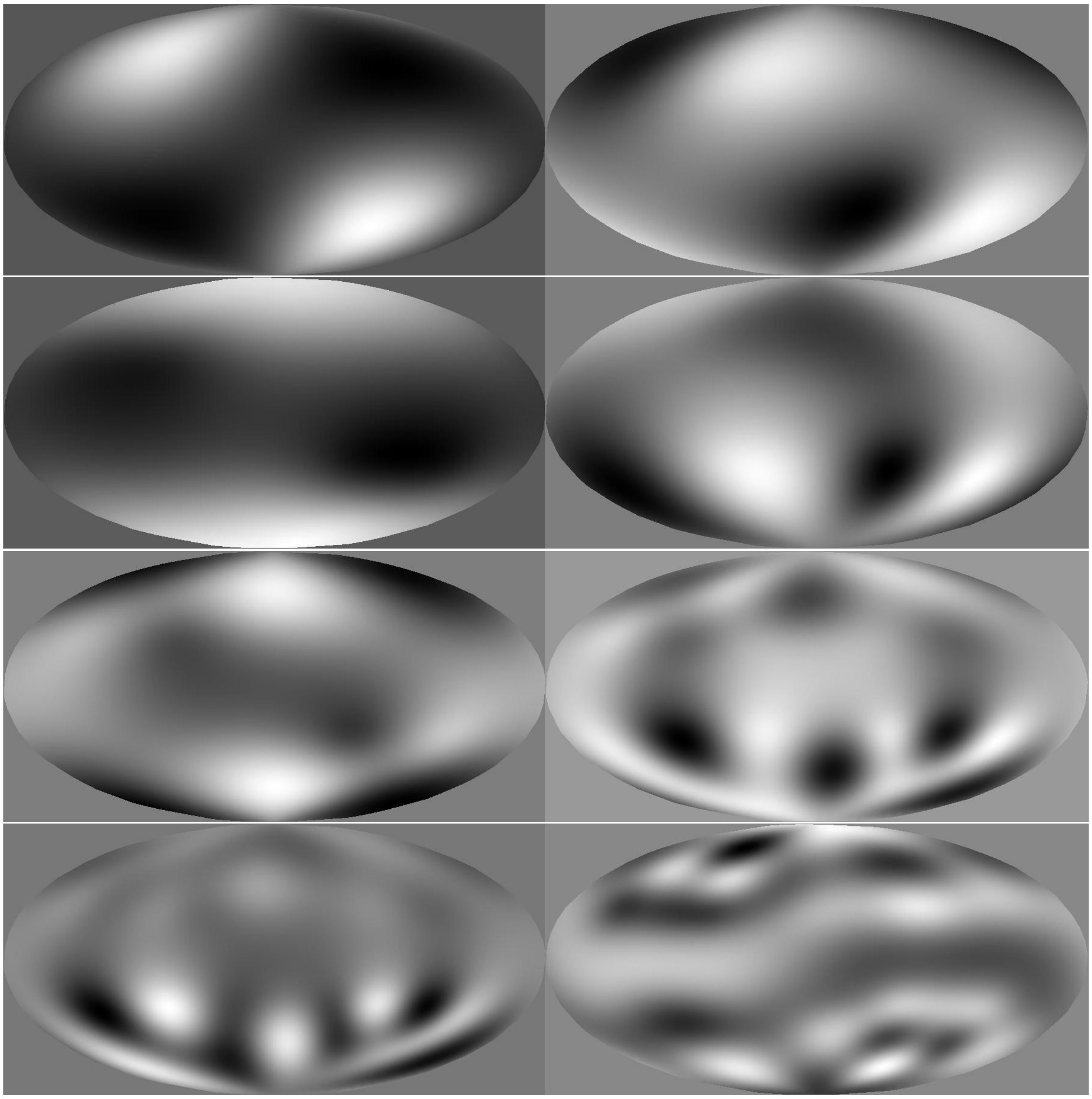

Figure 4.1: DMR signal-to-noise eigenmodes

A sample of the DMR eigenmodes are shown. As you read from left to right and top to bottom, the modes are sorted in order of decreasing eigenvalue. Modes 1, 2, 3, 4, 10, 20, 30, 40, 100, 200, 300, and 400 are plotted. This figure continues to the next page.

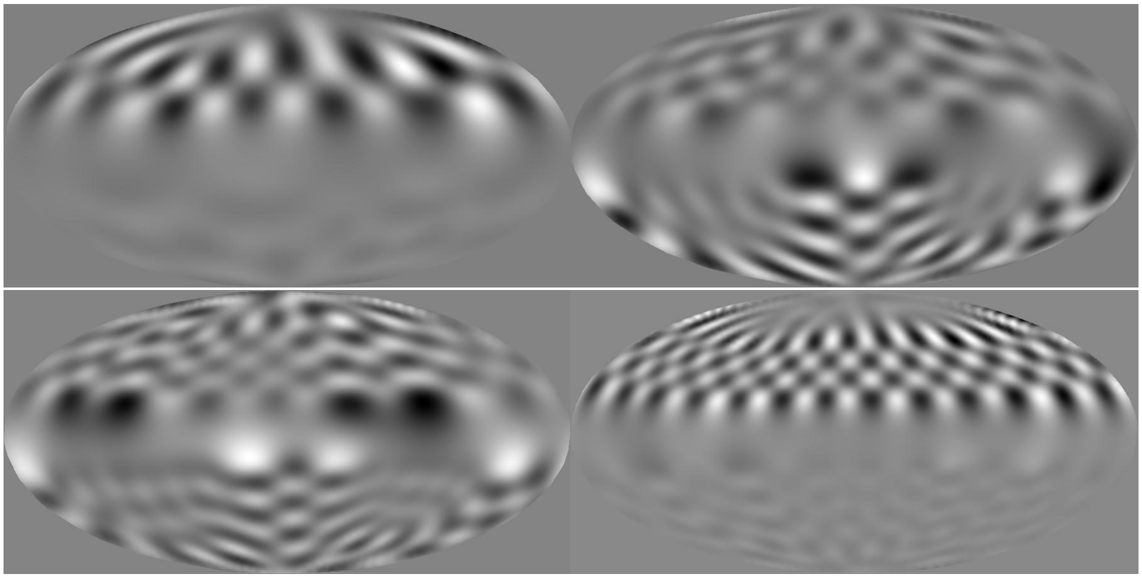

Figure 4.1 (continued)

mode 500 for $l_{\rm max} = 30$ and mode 1000 for $l_{\rm max} = 40$. We can test this by computing the difference between modes computed with different cutoffs. To be specific, let $\vec{\alpha}$ be an eigenmode computed with $l_{\rm max} = 30$ and let $\vec{\beta}$ be the corresponding eigenmode computed with $l_{\rm max} = 40$. Take both vectors to have magnitude 1. Then $\Delta = 1 - (\vec{\alpha}\cdot\vec{\beta})^2$, the square of the sine of the angle between the two vectors, is a natural measure of the difference between them. This quantity is typically less than 0.01 for the first 400 modes.[9]

We saw in equation (4.27) above that the sum of the squares of eigenvalues is a measure of the expected sharpness of the peak of the likelihood function, and that we can therefore throw away all modes that do not make a significant contribution to the sum. In Figure 4.3 we show the eigenvalues λ_a as a function of mode number a, and also the running sum of the squares of the eigenvalues. It is clear from this plot that most of the power comes from the first 200 modes, but in order to be conservative we have chosen $K = 400$ as the number of modes to keep in our analysis. Figure 4.4 shows likelihood curves for a standard Sachs-Wolfe spectrum as a function of the normalization $Q_{\rm ps}$ for various values of K. As expected, the likelihood curve does not get appreciably narrower for $K > 200$.

---

[9] Occasionally a near degeneracy between two modes will be broken in different ways for different $l_{\rm max}$'s. The parameter Δ will then be large. This phenomenon, which was observed once (at mode 400) in a spot check of 20 modes, is completely benign.

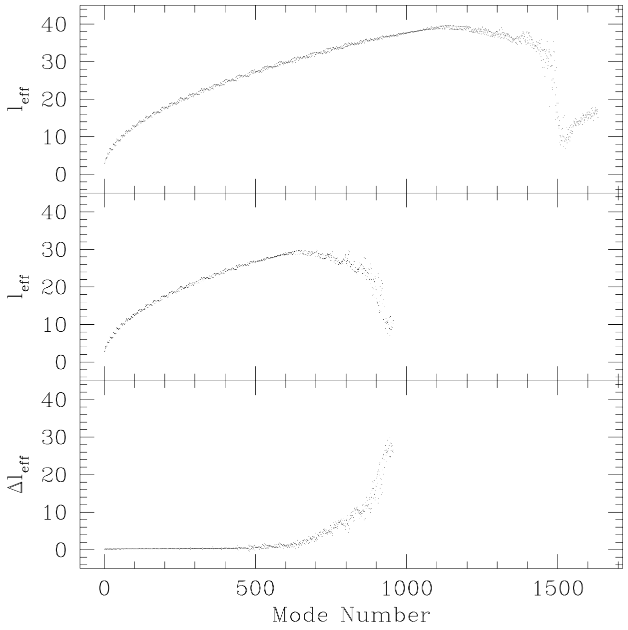

Figure 4.2: Effective l-values of eigenmodes

We plot the quantity $l_{\rm eff}$ defined in the text. The points plotted in the top panel were computed using an $l_{\rm max}$ of 40, while those in the middle panel were computed with $l_{\rm max} = 30$. The difference between the two is plotted in the lower panel.

4.4 Preliminary Results

Having computed the eigenmodes, we are in a position to start computing likelihoods for every model we can think of. We will describe several such batches of models in the next chapter. For the moment we will content ourselves with mentioning some preliminary results, most of which pertain to the simple Sachs-Wolfe family of power spectra described by equation (2.29). We therefore have only two parameters to adjust, the normalization $Q_{\rm ps}$ and the slope n.

If we restrict our attention to the Harrison-Zel'dovich $n = 1$ spectrum, then the likelihood function $L(Q_{\rm ps})$ is approximately Gaussian, with peak value 21.1 μK and standard deviation 1.54 μK. For other values of n the normalization is

$$Q_{\rm ps} = 21.1 \exp\left[0.69(1-n)\right]\ \mu{\rm K} \tag{4.36}$$

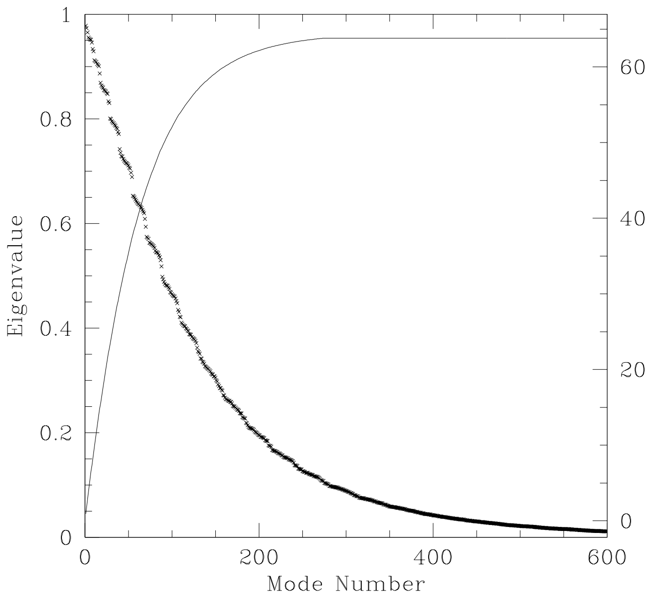

Figure 4.3: Signal-to-noise eigenvalues

The crosses mark the eigenvalues, which have been sorted in decreasing order. The solid curve is a running sum of the squares of the eigenvalues, with the scale marked on the right of the plot.

for $0.5 < n < 2$. The maximum-likelihood point is $(Q_{\rm ps}, n) = (16.8\,\mu{\rm K}, 1.32)$; marginalizing over the quadrupole changes this to $(18.6\,\mu{\rm K}, 1.2)$. The 68% credible range is $0.93 < n < 1.54$ ($0.72 < n < 1.40$ without the quadrupole). These credible regions are obtained by marginalizing over the normalization. We chose to marginalize with a prior that is uniform in $Q_{\rm ps}$; another reasonable choice is to choose a prior uniform in the normalization at the "pivot point" C_9. (The pivot point is the value of l such that C_l is roughly unchanged as we tilt the spectral index n.) The results obtained with these priors are negligibly different.

We need not restrict ourselves to a simple two-parameter family of power spectra. If we are willing to overwork our computer, we can fit to as large a parameter space as we wish. In fact, we can take *all* of the C_l's over some range of l to be free parameters. Armed with *Numerical Recipes* [143], we can search this whole parameter space for the point where the likelihood is maximized.

The result of this exercise in overzealous fitting is shown in Figure 4.5. The power

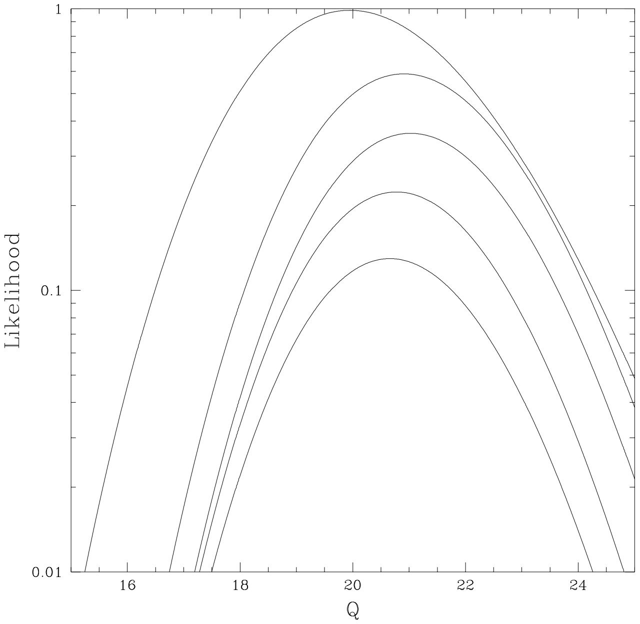

Figure 4.4: The effect of varying the number of modes

We plot the likelihood curve $L(Q_{\rm ps})$ for a Harrison-Zel'dovich Sachs-Wolfe spectrum for various values of the number K of modes included. The normalizations are arbitrary. The values of K for the various curves are 100, 200, 300, 400, and 500 from the top of the plot to the bottom.

spectrum shown was obtained by maximizing the likelihood over all C_l with $2 \leq l \leq 19$. Modes with $l > 19$ were held fixed at the same values as the Harrison-Zel'dovich spectrum, although the results are quite insensitive to varying this choice. The likelihood for this model is 100 times that of the best-fitting $n = 1$ model.

The error bars on the points result from interpreting $L(C)$ as a probability distribution for C (*i.e.*, assuming a uniform prior distribution in C_l) and approximating L as a multivariate Gaussian near the peak. The error bars are just the square roots of the diagonal elements of the covariance matrix of this distribution. In general off-diagonal elements of this matrix are *not* negligible. In particular, the estimates are correlated with their second-nearest neighbors.[10] The dimensionless correlation $\mathrm{Cov}(C_l, C_{l'})/\sqrt{\mathrm{Cov}(C_l, C_l)\mathrm{Cov}(C_{l'} C_{l'})}$

---

[10] Nearest-neighbor correlations are strongly suppressed by the nearly perfect reflection symmetry of our problem. Even and odd modes are almost completely uncoupled.

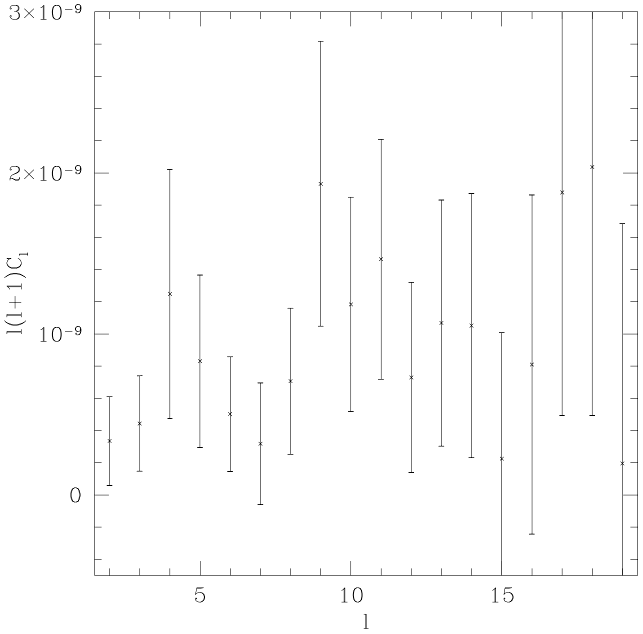

Figure 4.5: The maximum-likelihood power spectrum

The power spectrum shown here was determined by maximizing the likelihood taking C_l for $2 \leq l \leq 19$ as free parameters. The error bars were computed by interpreting $L(C)$ as a probability distribution for C_l and approximating L as a Gaussian near the peak. There are non-negligible correlations between the errors.

ranges from about 0.2 at low l to 0.05 at high l for second-nearest-neighbor pairs. In most cases, correlations between points with $|l - l'| \neq 2$ are negligible.

However, even knowing the full covariance matrix does not make the data in Figure 4.5 particularly useful for assessing other models. The reason is that the Gaussian approximation for $L(C)$ is quite poor. In particular, the likelihood is very strongly skew-positive, meaning that power spectra that lie slightly above the data points in the figure fit better than those that lie slightly below them. For example, if the best-fitting $n = 1$ model were plotted in Figure 4.5, it would be a horizontal line at $l(l+1)C_l = 9.0 \times 10^{-10}$, which is noticeably higher than one would get by naively fitting a horizontal line to the points in the figure.

4.5 Tests of the Karhunen-Loève Method

The first thing we should probably do is assess the consistency of our results with those of other methods. Tegmark & Bunn have performed a brute-force analysis based on the full likelihood function without expanding in any set of basis functions [174]. They restricted their attention to the two-dimensional space of Sachs-Wolfe power spectra, finding that the maximum-likelihood spectrum has parameters $(Q_{\rm ps}, n) = (18.2\,\mu{\rm K}, 1.15)$, or $(21.3\,\mu{\rm K}, 0.95)$ if the quadrupole is excluded. The 68% credible region for n is 1.10 ± 0.29. The discrepancies between these results and those described in the previous section are all much smaller than the uncertainties. In Figure 4.6 we plot the marginal likelihood for n from the brute-force analysis together with the Karhunen-Loève result and that of a similar analysis by Górski *et al.* [71]. To caution the reader, we also plot the likelihood function that one would obtain by neglecting to marginalize over the monopole and dipole. Although all the results agree fairly well, we should perhaps be concerned that the Karhunen-Loève estimate is slightly high compared to the others. It is therefore of interest to explore the possibility of bias in n with Monte Carlo simulations.

In order to test for bias in our parameter estimates, we must resort to Monte Carlo simulations. One of our primary goals is to estimate the normalization $Q_{\rm ps}$ for a variety of spectral shapes. We therefore performed simulations to see whether the maximum-likelihood estimate of $Q_{\rm ps}$ is unbiased. We made 1000 realizations of the CMB sky with each of three power spectra, an $n = 1$ Sachs-Wolfe spectrum, an $n = 1.5$ Sachs-Wolfe spectrum, and the "standard" cold dark matter spectrum described in the next chapter. In every case we chose an input normalization equal to the maximum-likelihood normalization of the real data. We added independent Gaussian noise with the same distribution as the actual DMR noise to our simulated skies and determined the maximum-likelihood value of $Q_{\rm ps}$ corresponding to the input power spectrum. In all three cases the mean of these estimates differed from the input normalization by less than $0.1\,\mu$K, and the standard deviation of the estimated normalizations was about 8% of the mean.

We would also like to know whether the maximum-likelihood value of n is an unbiased estimator. In order to test this, we created 315 simulated skies with an $n = 1$ power spectrum, added noise, and found the maximum-likelihood point in the $(Q_{\rm ps}, n)$ plane. The mean value of n from these trials was 1.04, and the standard deviation was 0.30. We take these results to indicate that any bias in our estimate of n is fairly small. These simulations

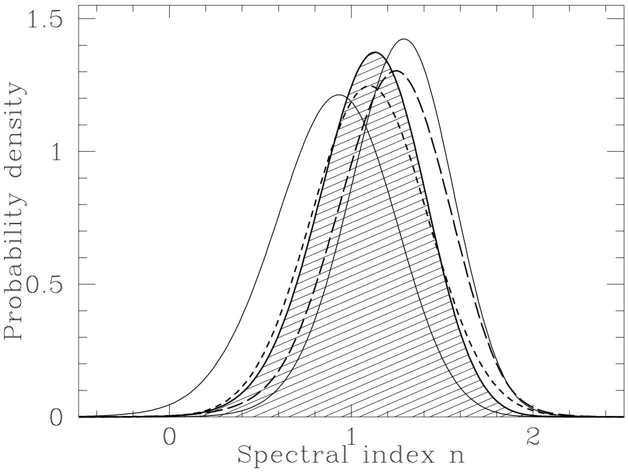

Figure 4.6: Marginal likelihoods for n with different methods

In order of decreasing peak height, the curves are obtained by the naive brute force method (neglecting to account properly for monopole and dipole removal), the brute force method (shaded), the Karhunen-Loève method described in this chapter (dashed), Górski *et al.* [71] (dotted) and the brute force method with quadrupole removed, respectively. All five curves are based on the combined 53 and 90 GHz two year COBE data and are marginalized over the normalization at the "pivot point" C_9. All but the last curve include the quadrupole.

probe a somewhat atypical case, in that the input power spectrum for the simulations is identical to the fiducial power spectrum. We therefore repeated the test with an $n = 2$ input power spectrum. After 275 simulations, the average recovered value of n was 1.98, and the standard deviation was 0.24. We can therefore be reasonably confident that our n estimate is unbiased.

We can also use these simulations to define frequentist confidence intervals for n. In 68% of our simulations, n differed by its true value by less than 0.29, and so we estimate that $n = 1.32 \pm 0.29$ at 68% confidence. The corresponding error bars for 90% and 95% confidence are ± 0.50 and ± 0.58 respectively. Of course, given the relatively small number of simulations we performed, these limits are only approximate. Furthermore, they are only strictly valid for the $n = 1$ power spectrum that was used in the simulations.

It is clearly of interest to check whether the likelihoods obtained by the Karhunen-Loève method are sensitive to the choice of fiducial power spectrum. Our working hypothesis is that the transformed data contain most of the useful cosmological information and that the modes that were thrown away are dominated by noise. If this is true, then we should expect our results to be insensitive to variations in the input power spectrum.

In order to test this hypothesis, we repeated some of the analysis described in the previous section, replacing the $n = 1$ Sachs-Wolfe spectrum, $C_l^{-1} \propto l(l+1)$, with the $n = 2$ spectrum,

$$C_l \propto \frac{1}{l+\frac{1}{2}}. \tag{4.37}$$

The difference between these two spectra is fairly extreme, and so we expect any significant sensitivity to the choice of fiducial power spectrum to show up in a comparison of the two.

The likelihood function based on the $n = 2$ power spectrum peaks at the point $(Q_{\rm ps}, n) = (16.7\,\mu{\rm K}, 1.33)$, differing by only $0.1\,\mu$K in normalization and 0.01 in n. The Harrison-Zel'dovich normalization changes from $21.1\,\mu$K to $20.9\,\mu$K when we change the fiducial power spectrum.

Another quantity of interest in assessing sensitivity to the fiducial power spectrum is the likelihood ratio of models with different slopes. Let us define r to be the ratios of peak likelihoods at two different slopes n and n':

$$r(n,n') = \frac{\max_{Q_{\rm ps}} L(n, Q_{\rm ps})}{\max_{Q_{\rm ps}} L(n', Q_{\rm ps})}. \tag{4.38}$$

We would like to know whether r changes when we change the fiducial power spectrum. As we change from an $n = 1$ input power spectrum to one with $n = 2$, $r(1.25, 1)$ changes from 1.69 to 1.74, and $r(1, 0.5)$ changes from 15.6 to 18.8. These changes are fairly small, and give us confidence that our Karhunen-Loève results are not strongly sensitive to the choice of fiducial power spectrum.

We conclude this section with an important note about normalizations. Our results are consistently about 5% higher than those obtained in a likelihood analysis done by the COBE team [71]. In particular, the COBE group finds a maximum-likelihood normalization for an $n = 1$ Sachs-Wolfe spectrum of $20\,\mu$K, while we find $21.1\,\mu$K. The COBE technique is somewhat different from our own, and the discrepancy is less than the one-σ uncertainty in either result, but, nonetheless, one might have expected closer agreement. It has been suggested [69] that the discrepancy may be due to the fact that we use the ecliptic maps rather than the Galactic maps used by the COBE group: when the COBE team's analysis is performed on the ecliptic map, the discrepancy is reduced to less than $0.5\,\mu$K [73].

We confirmed this hypothesis by repeating our analysis on the Galactic sky maps, using exactly the same techniques as in our original analysis. Using the Galactic data, the maximum-likelihood normalization for an $n = 1$ Sachs-Wolfe spectrum is $Q_{\rm ps} = 20.1\,\mu{\rm K}$, which agrees almost exactly with the COBE result. The maximum-likelihood point in the $(Q_{\rm ps}, n)$ plane based on this data set is $(15.9\,\mu{\rm K}, 1.35)$, differing only very slightly in slope from the ecliptic result. We know of no reason to prefer one data set over the other, and we therefore conclude that there is some error intrinsic to the map-making process that affects normalizations at a level of $\sim 5\%$. Furthermore, the primary difference between the two maps would seem to be the normalization: shape parameters such as n seem to be less sensitive to the difference.

4.6 Placing Frequentist Constraints on Models

Statistical methods for using a data set like COBE to place constraints on models come in two general varieties, Bayesian and frequentist. Most CMB work, including analyses of the COBE data as well as other experiments, has taken a Bayesian point of view, and in this dissertation we will generally follow this trend. However, in this section we will present a simple technique for using the Karhunen-Loève-transformed data to place frequentist constraints on models.

Once we have chosen a prior distribution, a Bayesian approach is perfectly adequate for assessing the *relative* merits of the various models under consideration: with this approach we can say, for example, that model A is 10 times more likely than model B.[11] These models may only differ by a normalization or could be drawn from different cosmologies or structure-formation scenarios. In some cases, however, we would like to assign *absolute* consistency probabilities to models. The Bayesian approach is not well suited to answering this sort of question, and the problem of assigning an absolute consistency probability to a model is best attacked with frequentist methods. Consider for example the definition of a Bayesian credible region in equation (4.4). It singles out the regions in a particular parameter space that have the highest likelihood relative to the rest of parameter, but it makes no attempt at all to assess whether any of the parameter space is an intrinsically good fit.

In the frequentist approach, we choose some goodness-of-fit statistic η, and compute

---

[11] The choice of prior distribution can of course be quite vexing. Most of results we give in this dissertation are relatively insensitive to this choice, as long as we restrict our attention to fairly broad, "uninformative" priors.

its probability distribution over a hypothetical ensemble of realizations of the model. We then compute the value of η corresponding to the real data, and determine the probability P of finding a value of η as extreme as the observed value in a random member of our ensemble. We take this probability to be a measure of the consistency of the data with the model: if the such an extreme value of η does not occur often in realizations of the model, we say that the model can be ruled out at a significance level P.

Two points about this technique deserve emphasis. First, the significance levels derived in this manner are conceptually quite distinct from Bayesian likelihoods. Bayesians and frequentists ask different questions of their data, and will therefore sometimes get different answers. We do expect that models that have low Bayesian likelihoods will in general have low frequentist P-values; however, there is no generally applicable quantitative relation between the two. Second, it is clear that the P-values in the frequentist approach depend on the goodness-of-fit statistic η. For some classes of problems a standard choice is available; for the problem we consider below, we are unaware of such a standard.

The first goodness-of-fit statistic one might think of for this purpose is a simple chi-squared,

$$\chi^2 \equiv \sum_{i=1}^{K} \left(\frac{x_i}{\sigma_i} \right)^2 , \tag{4.39}$$

where x_i is the amplitude of the ith element in our eigenmode expansion and σ_i^2 is the variance predicted for x_i by our model.[12] (In order to remove all sensitivity to the monopole and dipole, the eigenmodes α_i should be orthogonalized with respect to these modes before the x_i are computed.) This statistic would be a natural choice if we wished to constrain the normalization of a model; however, our primary interest here will be in constraining the *shape* of the power spectrum, and this goodness-of-fit statistic is not well suited for this purpose. In fact, given *any* power spectrum C_l, we can choose a normalization that gives a χ^2 that lies exactly at the median of its probability distribution, since σ_i scales with the normalization of the theory. We would therefore conclude that for some normalization this model is a perfectly good fit regardless of the shape of the C_l.

To focus on the power spectrum, let us consider quantities quadratic in the amplitude of the eigenmodes. We could for example define a χ^2-like statistic based on x_i^2 instead of

---

[12] Since the x_i are not independent, this statistic is not chi-squared distributed. We would probably have to estimate its distribution by Monte Carlo simulation.

x_i:

$$\chi^2 = \sum_{i=1}^{K} \frac{(x_i^2 - \sigma_i^2)^2}{2\sigma_i^4}. \tag{4.40}$$

If our model is correct, then each x_i^2 is chi-squared distributed with mean σ_i^2. The variance of such a distribution is $2\sigma_i^4$, and so we expect to get a contribution of about one per degree of freedom from this sum if our model is correct. Furthermore, we expect the value of this statistic to be higher for incorrect models than for the true model, so that it avoids the fatal flaw of the statistic in equation (4.39).

We can improve on this choice of statistic by binning the x_i^2 together. After all, as long as C_l is a reasonably smooth function, we expect any incorrect model to err in the same direction for all nearby modes. That is, if we have underestimated σ_i^2, we expect that we have probably underestimated σ_{i+1}^2 as well. (Recall from Figure 4.2 that nearby modes probe similar l-values.) We therefore expect binning to improve the power of our goodness-of-fit statistic.

Let us define

$$z_i = \sum_{j=(i-1)J+1}^{iJ} \left(\frac{x_j}{\sigma_j}\right)^2 \tag{4.41}$$

for $1 \le i \le K/J$. We should choose the bin size J to be large enough to reduce the intrinsic width of the distribution of z_i to a reasonable level, yet small enough that the mode amplitudes in each bin probe similar angular scales. We have adopted $J = 10$ as a compromise between these two considerations.

If our model is correct, then each z_i will be approximately J. If the model is incorrect, then some z_i will be too low, and others will be too high. For example, if our model has too little large-scale power, then the ratios x_j^2/σ_j^2 will be greater than 1 for small j, and the first few z_i will tend to be larger than J. We can quantify this observation by defining the goodness-of-fit statistic

$$\eta = \sum_{i=1}^{K/J} (z_i - J)^2. \tag{4.42}$$

Since the mode amplitudes x_i are in general correlated, it is not possible to compute analytically the probability distribution of η. We must therefore resort to Monte Carlo simulations. For each of several models, we created 1000 random sky maps. We added noise to each pixel by choosing independent Gaussian random numbers with zero mean and standard deviations corresponding to the noise levels in the real data. We computed

Model	D_1'	D_1''	L	P-value
1	0.4	-1.6	3.29	22.0%
2	0	0	1.00	13.7%
3	-0.5	3.5	0.091	7.0%
4	-2	0	0.014	2.8%
5	-0.342	2.455	0.186	7.6%
6	-0.275	1.561	0.310	8.6%

Table 4.1: Goodness of fit for six models

The shape parameters D_1' and D_1'' are defined in Section 5.3. L denotes the Bayesian peak likelihood, normalized so that a pure Harrison-Zel'dovich Sachs-Wolfe model has $L = 1$. The P-value is the percentage of simulated data sets for which the goodness-of-fit statistic η defined in equation 4.42 is greater than the value found for the real data.

the statistic η for each map. We chose to simulate six different models, which will be described in detail in Section 5.3 below. The first four were chosen to span a range of values of the Bayesian likelihood: we simulated (1) a Harrison-Zel'dovich spectrum; the models with the highest (2) and lowest (3) Bayesian likelihoods from the grid of quadratic power spectra described in the next chapter; and a model (4) with an even lower Bayesian likelihood $L = 0.01 L_{HZ}$. In addition, we chose two models from Table 5.2 that have identical cosmological parameters ($\Omega_0 = 0.1$, $h = 0.75$, $n = 0.85$), except that one has only scalar perturbations (5) and one includes tensors (6) in the ratio $C_2^T/C_2^S = 7(1-n)$. All of these models are approximated extremely well by quadratic power spectra as described in Section 5.3, and so their shapes can be completely characterized by the two parameters D_1' and D_1''. We give these parameters in Table 4.1.

Table 4.1 also shows the peak Bayesian likelihood for each model, as well as the frequentist P-value for the goodness-of-fit statistic η. It is clear from the Table that low Bayesian likelihoods tend to correspond to low P-values, as expected. Furthermore, those models with likelihoods of order unity relative to a pure Harrison-Zel'dovich spectrum fit the data reasonably well. This is a very reassuring fact: it was perfectly possible *a priori* that all of the models we have been considering would prove to fit the data poorly.

The z_i for a Harrison-Zel'dovich spectrum and our worst-fitting model (model 4 of Table 4.1) are shown in Figure 4.7. (In making Figure 4.7, we chose the coarser bin size

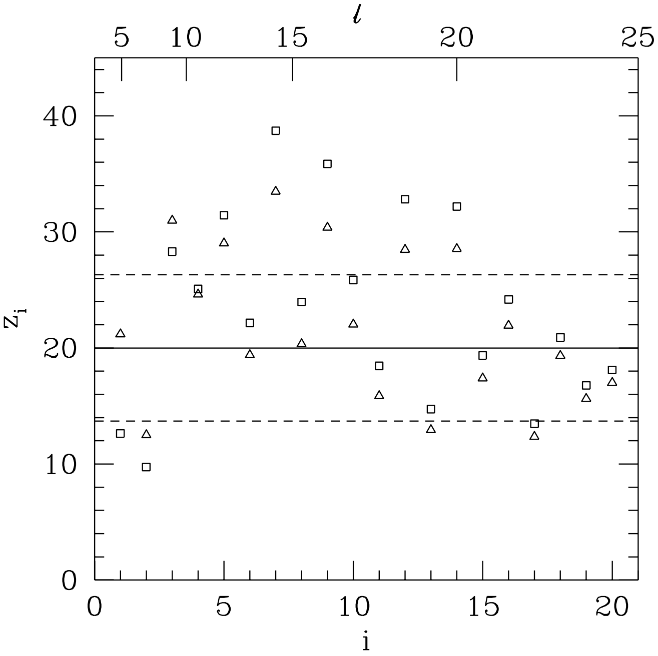

Figure 4.7: Measuring the goodness of fit

The quantities z_i defined in equation (4.41) are plotted. The triangles are the z_i computed with a Harrison-Zel'dovich spectrum and the squares were computed with our worst-fitting model, model 4 of Table 4.1. The z_i shown were computed with a bin size $J = 20$. The solid and dashed lines show the expected value of each z_i and approximate one-sigma deviations from it. (The z_i are only approximately chi-squared distributed and only approximately uncorrelated, so these estimated deviations are not precise.) The effective mode numbers l_{eff} are indicated at the top of the figure.

$J = 20$ rather than $J = 10$, to reduce scatter in the points.) Note that the z_i only loosely correspond to C_l and depend on the theory. When z_i is larger than its expected value, one can conclude that the data have more power than the theory predicts on the corresponding angular scale; however, there is *no* direct proportionality between z_i and the corresponding C_l. Each z_i can be regarded as an estimator of the power spectrum of the signal and noise *combined.* For small i, z_i samples mostly signal, while the noise dominates for large i. The value of z_i for large i therefore changes very little as the model parameters are varied, as can be seen in Figure 4.7.

We have adopted our goodness-of-fit statistic η in a fairly arbitrary manner. We certainly have no guarantee that it is a particularly powerful discriminator among models. It is well within the realm of possibility that a more clever choice of statistic would considerably strengthen our ability to place frequentist constraints on models. Nonetheless, this statistic is simple to compute, and gives us something that we can plot to get some idea of how well our data match a particular model over different ranges of l.

4.7 Conclusions

Almost all analyses of the DMR data have involved some form of "data compression," either by taking linear combinations of the data points or by forming quadratic or higher-order statistics. In this chapter we have presented a linear compression technique that is in some sense optimal. The technique involves one moderately computationally expensive task, namely, finding the signal-to-noise eigenmodes. However, the process of computing likelihoods after that is relatively efficient, involving nothing worse than the Cholesky decomposition of a 400×400 matrix.

Based on the tests performed in Section 4.5, we are reasonably confident that the Karhunen-Loève likelihood forms an adequate substitute for the much more computationally expensive "true" likelihood. Furthermore, we have performed Monte Carlo simulations that allow us to conclude that the maximum-likelihood normalization derived from the transformed data is unbiased. We also conclude that the bias in shape parameters, such as the spectral index n, appears to be small, even for spectral shapes that are somewhat different from the shape of the fiducial power spectrum.

The Karhunen-Loève-transformed data can be used to set frequentist constraints by choosing a goodness-of-fit statistic η in the usual way. However, there is a great deal of arbitrariness inherent in choosing this statistic. We have suggested one possible choice in the previous section, but we have no great confidence in its effectiveness, and we will therefore stick to likelihood-based Bayesian techniques for the rest of this dissertation.

Chapter 5

Constraining Models with the COBE Data

> "That process," said I, "starts upon the supposition that when you have eliminated all which is impossible, then whatever remains, however improbable, must be the truth."
>
> Arthur Conan Doyle

5.1 Introduction

Fictional characters have certain advantages not shared by the rest of us. Apparently, for Sherlock Holmes, the space of theories divides neatly into the possible and the impossible, with no inconvenient grey areas [54]. We are not so lucky. Even when we claim to have "ruled out" a particular theory, all we mean is that we have reached the conclusion that it is extremely implausible, not that it is impossible.

In this chapter we will attempt to place such constraints on several popular cosmological models, those based on the cold dark matter scenario. All of these models have at least one free parameter, the overall normalization $Q_{\rm ps}$ of the fluctuations, whose value can be accurately estimated from the COBE data. Having fixed the normalization, we can ask whether a particular model agrees with large-scale structure observations such as the fluctuations in galaxy counts and measurements of the velocity field.

One might imagine several different methods for estimating the normalization $Q_{\rm ps}$ from the COBE data. We have described in the previous chapter the one that we will be using. Despite its gain in efficiency over the "brute-force" technique of computing the full likelihood function, it is still a fairly cumbersome method, and one might well wonder why simpler techniques do not work. In particular, the expected pixel variance $\sigma^2(7^\circ)$ of the COBE data (after correction for noise bias) is proportional to $Q_{\rm ps}^2$:

$$\langle\sigma^2(7^\circ)\rangle = \frac{Q_{\rm ps}^2}{5}\sum_{l=2}^{\infty}(2l+1)\frac{C_l}{C_2}W_l^2, \tag{5.1}$$

where W_l is the Legendre-polynomial expansion of the beam pattern. (Recall that a particular theoretical model predicts the shape of the power spectrum, *i.e.*, C_l/C_2.) In order to reduce the noise, one sometimes smooths the data with a 7° FWHM Gaussian to get the variance on an angular scale of 10°:[1]

$$\langle\sigma^2(10^\circ)\rangle = \frac{Q_{\rm ps}^2}{5}\sum_{l=2}^{\infty}(2l+1)\frac{C_l}{C_2}W_l^2 e^{-\sigma_7^2 l(l+1)}, \tag{5.2}$$

where $\sigma_7 = 0.4247 \times 7^\circ$ is the standard deviation of the 7° Gaussian. Since both of these quantities have expected values proportional to $Q_{\rm ps}^2$, we might in principle choose to use either to set our normalization. Why don't we?

The reason is simply that reducing the data to a single number in either of these ways throws away a lot of information. Another way to see this is to observe that both $\sigma(7^\circ)$ and $\sigma(10^\circ)$ have large cosmic variances, for the simple reason that they are dominated by the first few terms in the sum, and therefore few independent modes of the anisotropy contribute. A likelihood technique of the sort we employ uses more of the information available in the data and therefore provides more precise normalizations.

The difference can be quite substantial. In Figure 5.1 we show the ensemble-averaged quantities $\langle\sigma^2(7^\circ)\rangle^{1/2}$ and $\langle\sigma^2(10^\circ)\rangle^{1/2}$ derived from simple Sachs-Wolfe power law models with the maximum-likelihood normalization. We can see that there are large differences between σ and $\langle\sigma^2\rangle^{1/2}$ in these models. The sense of the variation with n is as we would expect, given that the sky variance σ^2 is dominated by the quadrupole. For low n the quadrupole is much lower than its expected value, so σ is much lower than $\langle\sigma^2\rangle^{1/2}$. As n increases this effect decreases. In short, normalizing to a single number such as $\sigma^2(7^\circ)$ or $\sigma^2(10^\circ)$ is a very dangerous practice.

The fact that the locally-measured quantity $\sigma(10^\circ)$ differs noticeably from the ensemble-average quantity $\langle\sigma^2(10^\circ)\rangle^{1/2}$ (for, say, a maximum-likelihood $n = 1$ Sachs-Wolfe

---

[1] 10° being the the approximate quadrature sum of the original 7° FWHM beam-width and the additional 7° Gaussian.

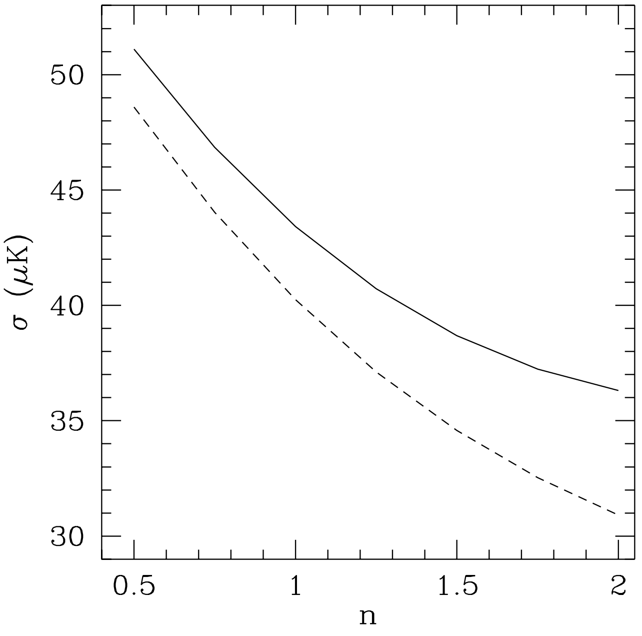

Figure 5.1: Expected pixel variance as a function of n

We show the value of the RMS power, as measured by $\sigma(7^\circ)$ (solid) and $\sigma(10^\circ)$ (dashed), for the best fitting, tilted Sachs-Wolfe models, as a function of n. The fact that the RMS fluctuation of the best fitting model depends on n shows that there is more information in the COBE data than just the RMS power. For comparison, the actual values of $\sigma(7^\circ)$ and $\sigma(10^\circ)$ in the COBE data are $37.5\,\mu$K and $32.5\,\mu$K respectively.

spectrum) seems to have caused a fair amount of consternation in the literature [6, 13, 31, 195], despite the fact that it is nothing more than a manifestation of cosmic variance. We will therefore comment briefly on the significance of, and the reasons for, the discrepancy.

One relevant fact is that the COBE beam is not a pure 7° Gaussian [195]. If we approximate W_l by a Gaussian in equation (5.2), we find that $\langle\sigma^2(10^\circ)\rangle^{1/2} = 1.99\,Q_{\rm ps}$ for a pure $n = 1$ model. Using the correct window function plus an extra 7° smoothing as in equation (5.2), one has $\langle\sigma^2\rangle^{1/2}/Q_{\rm ps} = 1.91$; the corresponding ratio for a standard CDM model is 1.95. These are to be compared with the value $\sigma/Q_{\rm ps} = 1.63$ obtained from the real data. Including the proper beam reduces the discrepancy only slightly.

But how significant is this discrepancy anyway? In a set of COBE-normalized standard CDM skies, we find that 7% have a ratio $\sigma/Q_{\rm ps}$ as low as that in the real data. Using

$n = 1.5$ CDM instead of standard CDM improves the situation: the predicted $\sigma/Q_{\rm ps}$ is 2.32, while the data give $\sigma/Q_{\rm ps} = 2.31$. In any case, even for $n = 1$ the discrepancy is not very statistically significant. The bulk of the discrepancy comes from the low COBE quadrupole. If we remove the quadrupole from the data, then $\sigma(10^\circ)$ and $Q_{\rm ps}$ are quite consistent with each other. The ratio found in the data after removing the quadrupole is $\sigma/Q_{\rm ps} = 1.54$, which agrees well with the prediction of 1.69 for sCDM.

We draw three conclusions from this analysis. First, the $\sigma/Q_{\rm ps}$ discrepancy is not terribly significant in any case. Second, the main reason for the discrepancy is simply the fact, which we have already observed, that the quadrupole in the data is low. Finally, we conclude that it is unwise to use a single number such as $\sigma(10^\circ)$ to normalize models.

As soon as we go beyond the minimal "standard" cold dark matter model, a host of new free parameters are introduced in addition to the normalization. Among the most important parameters we can vary are the spectral index n, the cosmological constant Λ, the contribution of tensor (gravitational wave) modes to the CMB anisotropy, the baryon fraction Ω_B, the amount of hot dark matter Ω_ν, and the Hubble constant. Some of these parameters (especially Λ) produce changes in the predicted spectral shape of the CMB anisotropy on COBE scales, and therefore change the normalization $Q_{\rm ps}$. Others, such as Ω_ν, have little effect on the COBE spectral shape but strongly affect the conversion from $Q_{\rm ps}$ to the amplitude of matter fluctuations on large-scale structure scales.

For those parameters that significantly affect the shape of the anisotropy spectrum, such as Λ and the tensor contribution C_2^T/C_2^S to the anisotropy, we can hope to use the COBE data to do more than just set the normalization. The likelihood techniques described in the previous chapter should have the power to discriminate good models from bad on the basis of the spectral shape.

Throughout this chapter we will use the Bayesian techniques and the Karhunen-Loève-transformed DMR data described in the previous chapter. In Section 5.2 we will estimate normalizations for a fairly small class of nearly standard CDM models. In section 5.3 we will broaden this class considerably by observing that nearly all CMB power spectra for flat CDM-like theories have shapes that are approximated well by a simple two-parameter family. We will provide estimated normalizations and relative likelihoods for a grid of models in this family. Finally, in Section 5.4 we will compute the likelihood as a function of the cosmological constant Λ in a cold dark matter model, and use it to place constraints on the value of the cosmological constant.

5.2 The COBE Normalization for Standard CDM

The CDM model (see *e.g.*, [112, 131]) has become the 'straw man' model of structure formation. In this model, $\Omega_0 = 1$, with a variable fraction Ω_B residing in baryons and the rest in massive (non-relativistic) dark matter. The initial density fluctuations are assumed to be Gaussian distributed, adiabatic, scalar density fluctuations with a Harrison-Zel'dovich spectrum on large scales, *i.e.*, $P(k) \propto k^n$ with $n = 1$.[2] The 'standard' CDM model (sCDM) has come to mean the one with $H_0 = 50\,\mathrm{km/s\,Mpc^{-1}}$, as a compromise between age and H_0 constraints, and with $\Omega_B = 0.05$, for consistency with the best central value from big-bang nucleosynthesis studies [159, 181].

Since this model is in such widespread use, it is important to know as accurately as possible the COBE normalization for it. We have already observed above that the simple RMS amplitude of the fluctuations (either $\sigma(7^\circ)$ or $\sigma(10^\circ)$) is not an adequate normalization. It is much better to use the normalization given in the previous section for an $n = 1$ Sachs-Wolfe spectrum. However, even this is not entirely right, since as Figure 5.2 shows, a CDM model is not fitted well by a simple Sachs-Wolfe model, even if we allow n to vary. (The fact that the CDM power spectra in Figure 5.2 do not coincide well with the Sachs-Wolfe spectra indicates that other physical effects besides the simple Sachs-Wolfe effect significantly influence the angular power spectrum.) In this section we therefore provide COBE normalizations for the CDM power spectrum itself as computed by solving the Boltzmann equation for the matter-radiation fluid.

Although generally successful, the sCDM model is unlikely to be precisely the correct description of the Universe. In this section we will therefore consider a modest range of deviations from it. (We will widen the scope of our inquiry later in the chapter.) First one can allow for variations in h and Ω_B. On the scales probed by COBE, the power spectrum is relatively insensitive to these changes, although the conversion from the COBE normalization to measures of power on smaller scales can change dramatically as these parameters are varied. Second, since inflation generically predicts departures from the simple 'flat' $n = 1$ form, we will quote normalizations for a range of n. Should primordial gravitational waves (tensors) turn out to be important in addition to the scalar fluctuations, their effect will be to lower the inferred $Q_{\rm ps}^2$ by roughly the fraction they contribute to C_2,

---

[2] Inflation generally predicts slight deviations from the simple Harrison-Zel'dovich form. We neglect these corrections.

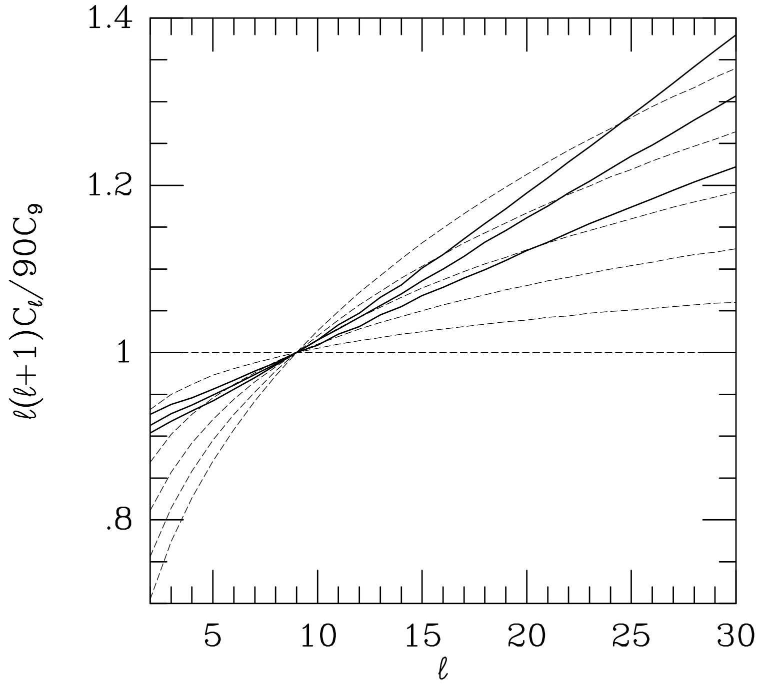

Figure 5.2: CDM and Sachs-Wolfe power spectra

The solid lines are the CMB anisotropy multipoles for CDM models for $\Omega_B = 0.01, 0.05$ ('sCDM') and 0.10 (from bottom to top). The dashed lines are Sachs-Wolfe spectra for $n = 1.00$ to $n = 1.25$ in steps of 0.05, as computed using equation (2.29). The curves have all been normalized at C_9, which is approximately the 'pivot' point for the COBE data [71]. Note that the CDM spectra are not represented well by Sachs-Wolfe spectra for *any* n.

i.e., C_2^T/C_2^S. This fraction is currently totally unknown, although it may be related to n for specific models. For example, in 'extended' inflation it is roughly $7(1-n)$ [44], although in 'natural' inflation it is negligible even if $n < 1$ [4]. Since including a tensor component also changes the shape of the C_l's, we perform a fit to the data. The two specific possibilities mentioned above span the usually considered range, and we quote the normalization in both cases. The addition of a component of hot dark matter will cause a negligible effect at COBE scales, but will change the relevant normalization on smaller scales. Models that have a cosmological constant and/or are open (including Baryonic Dark Matter models), and models containing topological defects, present additional difficulties, and we do not

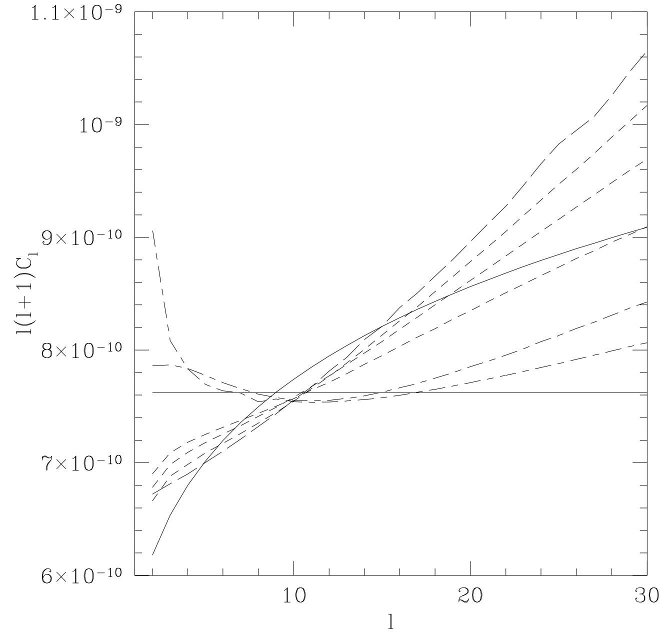

Figure 5.3: Power spectra for CDM models

We plot the angular power spectra for the family of CDM-like models whose normalizations are given in Table 5.1. The solid curves are pure Sachs-Wolfe spectra with $n = 1$ (flat) and $n = 1.15$ (rising). The short-dashed curves are the spectra for standard CDM and the models with varying Ω_B. (Models with larger values of Ω_B have more power at high l.) The spectrum for the $h = 0.3$ model is plotted with a long dashed line. The two tilted spectra are plotted with alternating long and short dashes. The model that includes tensors is the one with the higher quadrupole.

treat them in this section. We will treat cosmological-constant models later in the chapter; the reader with an interest in open inflationary models is referred to theoretical work by Lyth & Stewart [121], Ratra & Peebles [144], Bucher *et al.* [29] and Yamamoto *et al.* [197]. Some of these models have been compared with the DMR data by Górski *et al.* [72].

We have applied the likelihood techniques described in the previous section to the Karhunen-Loève-transformed two-year DMR data to determine the best-fitting values of $Q_{\rm ps}$ for this family of models. We used theoretical power spectra that were computed by Martin White. The results are shown in Table 5.1. The angular power spectra for all of these models are plotted in Figure 5.3

Model	$10^{11}A$	$10^{-5}B$ $(h^{-1}\mathrm{Mpc})^{3+n}$	$10^{10}C_2$	Q_{ps} (μK)	$\langle\sigma^2(10^\circ)\rangle^{1/2}$ (μK)	V_{60} (km/s)	V_{40} (km/s)	$J_3(10)$ $(h^{-1}\mathrm{Mpc})^3$	σ_8
S-W n=1	3.04	7.37	1.27	19.2	38.3	—	—	—	—
S-W n=1.15	2.06	18.38	1.03	17.3	36.5	—	—	—	—
sCDM	2.70	6.54	1.13	18.1	37.0	328	411	484	1.24
$\Omega_B = 1\%$	2.76	6.68	1.15	18.3	37.2	332	416	554	1.33
$\Omega_B = 10\%$	2.66	6.44	1.11	18.0	36.8	315	392	380	1.10
$h = 0.3$	2.68	6.49	1.12	18.1	35.0	274	329	128	0.62
tCDM (S)	3.49	3.55	1.31	19.5	37.4	278	343	280	0.94
tCDM (S+T)	2.38	2.42	1.51	21.0	39.8	230	283	191	0.77
MDM	2.70	6.54	1.13	18.1	37.0	332	415	264	0.87

Table 5.1: Normalizations for standard CDM models

The first three normalizations are proportional to Q_{ps}^2 and have fractional uncertainties $\simeq$ 15%, while the others are proportional to Q_{ps} and have errors $\simeq$ 7.5%. The first two models were normalized using equation (2.29) for the angular power spectrum, neglecting everything but the Sachs-Wolfe effect. The third model is sCDM. The next three models show the effect of varying the baryon fraction and the Hubble constant. tCDM is a tilted CDM model, with $n = 0.9$, given for the case with scalars only (S) and with a component of tensors added (S+T). MDM is a mixed dark matter model containing 30% hot and 69% cold dark matter, taken from [84]. The power normalizations A and B are defined in equations (5.3) and (5.5) respectively. Here by $\sigma^2(10^\circ)$ we mean the variance in ΔT smoothed with a 10° FWHM Gaussian. V_{60} and V_{40} are the predicted RMS velocities in spheres of radius $60h^{-1}$Mpc and $40h^{-1}$Mpc after smoothing with a Gaussian of width $\sigma = 12h^{-1}$Mpc. The $J_3(10)$ and σ_8 normalizations are defined in the text. The power spectra used to compute these results are plotted in Figure 5.3.

For sCDM we find $Q_{\rm ps} = 19.9 \pm 1.5\,\mu$K. If we unwisely marginalize over the quadrupole, this value increases slightly to $Q_{\rm ps} = 20.3 \pm 1.5\,\mu$K. Note that the effect of quadrupole removal on the normalization is quite small: it is considerably less than the 7.5% uncertainty in $Q_{\rm ps}$, and it is also less than the $\sim 5\%$ difference in normalization between the two pixelizations of the data (see Section 4.5). The reason throwing out the quadrupole does not affect the normalization much, despite the fact that Q is quite low, is that the quadrupole has large cosmic variance and is therefore automatically given low weight in fitting.

The above normalizations are mean values of $Q_{\rm ps}$ obtained by interpreting $L(Q_{\rm ps})$ as a probability distribution for $Q_{\rm ps}$. (In other words, by adopting a prior uniform in $Q_{\rm ps}$.) Since $L(Q_{\rm ps})$ is quite symmetric, they differ negligibly (by less than $0.1\,\mu$K) from the maximum-likelihood normalizations. If we choose to regard $Q^2_{\rm ps}$ as the quantity of interest and compute its mean with the same probability distribution, we get $Q^2_{\rm ps} = 403 \pm 61\,\mu{\rm K}^2$, and $Q^2_{\rm ps} = 418 \pm 64\,\mu{\rm K}^2$ without the quadrupole. These are slightly higher than the maximum-likelihood values. The value of $Q_{\rm ps}$ in CDM models depends weakly on Ω_B and is virtually independent of h, as indicated in Table 5.1; however, both Ω_B and h affect the transfer function and hence the smaller-scale normalizations. Including tensor modes degrades the fit to the data, since it increases the predicted quadrupole. With $C_2^T/C_2^S = 7(1-n)$ we find that the best-fitting values of $Q_{\rm ps} = (Q^2_{{\rm ps},T} + Q^2_{{\rm ps},S})^{1/2}$ are 22.9, 25.2 and $27.5\,\mu$K for CDM models with $n = 0.9$, 0.8 and 0.7, respectively. These models have peak likelihoods that are 0.61, 0.35, and 0.20 times the peak likelihood for the sCDM models. (The corresponding numbers for tilted CDM models without tensors are 0.75, 0.52, and 0.33.)

If we assume that temperature fluctuations at large scales arise from the SW effect, $\Delta T/T = {}^1\!/_3\Phi$, the spectrum of radiation fluctuations looks like

$$\mathcal{P}(k) \equiv A\,(k\eta_0)^{n-1}\,. \tag{5.3}$$

Here $\eta_0 \simeq 3t_0 = 2H_0^{-1}$ (for $\Omega_0 = 1$) is the conformal time today, with scale factor normalized to unity, and we are working in the synchronous gauge. A is one way of quoting the amplitude for scalar perturbations. It is related to ε_H, the dimensionless amplitude of matter fluctuations at horizon crossing, through $\varepsilon_H^2 = (4/\pi)A$.

In terms of the normalization A, the Sachs-Wolfe formula (2.29) becomes

$$C_l = 2^n\pi^2 A\,\frac{\Gamma(3-n)\Gamma\left(l+\frac{n-1}{2}\right)}{\Gamma^2\left(\frac{4-n}{2}\right)\Gamma\left(l+\frac{5-n}{2}\right)} \tag{5.4}$$

(see *e.g.*, [2, 25, 137, 188]). For the special case $n = 1$, we have $C_2/A = 4\pi/3$, and $C_l^{-1} \propto l(l+1)$. However, as has been emphasized before [19] and as we show in Figure 5.2, the C_l's for sCDM can depart significantly from equation (5.4) even for $l \sim 10$.

Another common normalization convention is to define the *matter* power spectrum as

$$P(k) \equiv Bk^n T^2(k), \tag{5.5}$$

where the transfer function $T(k) \simeq 1$ on large scales. The dimensions of B will thus depend on n (and will be Length^4 for $n = 1$). If the fluctuations arise purely from the SW effect, then B and A are simply related, so for $n = 1$ [188]

$$P(k) = 2\pi^2\eta_0^4\, A\, k\, T_m^2(k) \simeq 2.5 \times 10^{16} A\, (k/h\,\mathrm{Mpc}^{-1})\, T^2(k)\ (h^{-1}\mathrm{Mpc})^3. \tag{5.6}$$

Hence for $n = 1$,

$$B = 2\pi^2\eta_0^4 A = (6\pi^2/5)\,\eta_0^4\, Q_{\mathrm{ps}}^2/T_0^2. \tag{5.7}$$

The assumption of pure SW fluctuations is not exact (see Figure 5.2), although when we normalize to the quadrupole it is approximately correct. A calculation in which the photon and dark-matter perturbations are explicitly evolved in time shows this relation holds to an accuracy of $\lesssim 4\%$. There is better agreement for higher values of h, which moves last scattering further into the matter dominated regime. For example, if $h = 1$, the Sachs-Wolfe formula (for C_2) is good to $\lesssim 1\%$. We have used the ratio of matter to radiation normalizations from this calculation where appropriate. Additionally we use numerical calculations of the transfer functions for our models. We find that variations in the transfer function with Ω_B can lead to 10% changes in small-scale power.

Large-scale flows also provide a measure of the power spectrum: the variance of the velocity field in spheres of radius r, $V_{\mathrm{rms}}^2(r)$, can be expressed as an integral over the power spectrum (*e.g.*, [138]). This tends to probe scales similar to the degree-scale CMB experiments. Bertschinger *et al.* [16] estimated the 3D velocity dispersion of galaxies within spheres of radius $40h^{-1}$Mpc and $60h^{-1}$Mpc, after smoothing with a Gaussian filter on $12h^{-1}$Mpc scales. So we also quote the normalization in terms of the quantities V_{40} and V_{60}, corresponding to the procedure above.

On smaller scales, associated with clusters of galaxies, one conventionally quotes $J_3(10h^{-1}\mathrm{Mpc})$ (*e.g.* [138]), and the 'bias' b_ρ or variance of the density field in spheres of

$8h^{-1}$Mpc radius σ_8, defined through

$$b_\rho^{-2} \equiv \sigma_8^2 \equiv \int_0^\infty \frac{dk}{k}\ A(k\eta_0)^{n+3}\ T^2(k)\ \left(\frac{3j_1(kr)}{kr}\right)^2 , \tag{5.8}$$

where $r = 8\,h^{-1}$Mpc. The variance of galaxies, possibly biased relative to the matter ($\delta_{\rm gal} = b\delta_\rho$), is roughly unity on a scale of $8h^{-1}$Mpc [46, 120]. For sCDM the COBE best-fit gives $\sigma_8 \simeq 1.3$, *i.e.*, a slightly anti-biased model. However, as indicated in Table 5.1, σ_8 depends strongly on the values of Ω_B, h, etc.

5.3 Normalizations and Likelihoods for General Models

In the previous section we presented the normalization of the standard CDM model and a small range of variants. In this section we extend this to a larger class of models, and present a means for normalizing a whole class of models to the COBE data in a computationally simple manner. In particular, we will consider models with a cosmological constant and spectral tilt, both with and without the presence of tensor modes. Since this is a fairly large parameter space, we will simplify our computations by observing that all of the models under consideration are fitted extremely well by a simple two-parameter family of power spectra.

Since the power spectra for all of these models are quite smooth, it makes sense to consider approximating them with a Taylor expansion. Following [184], we will write

$$D(x) = l(l+1)C_l \qquad \text{with } x = \log_{10} l \quad . \tag{5.9}$$

We can perform a Taylor expansion of $D(x)$ about some fiducial point, which we shall take to be $x = 1$ ($l = 10$). Many theories (see below) can be approximated well by quadratic $D(x)$ over the relevant range for COBE, roughly $l = 2$ to 30, and so we present the normalizations and likelihoods of quadratic $D(x)$. We choose to parameterize our quadratics by the (normalized) first and second derivatives at $x = 1$: D_1' and D_1'', where

$$D(x) \simeq D_1 \left(1 + D_1'(x-1) + \frac{D_1''}{2}(x-1)^2\right) \tag{5.10}$$

(note that D_1' is $1/D_1$ times the derivative of $D(x)$ at $x = 1$). The normalization is then given by quoting D_1, or C_{10}, for each (D_1', D_1'') pair, and the goodness of fit is quantified by the relative likelihood of that shape compared to a featureless $n = 1$ Sachs-Wolfe spectrum.

Now to find the normalization of any theory, one calculates the large-angle multipole moments and finds the quadratic which best describes their shape. We have computed maximum-likelihood normalizations for these quadratic power spectra for a grid of models spanning the range $-0.5 \leq D_1' \leq 0.5$ and $-0.5 \leq D_1'' \leq 3.5$. The best-fitting amplitude and likelihood are given by the following formulae:

$$10^{11} C_{10} = 0.8073 + 0.0395 D_1' - 0.0193 D_1'', \qquad \text{and} \tag{5.11}$$

$$\ln L = 0.00697 + 1.523 D_1' - 0.403 D_1'^2 - 0.490 D_1'' - 0.0391 D_1' D_1'' + 0.00855 D_1''^2. \tag{5.12}$$

The fitting formula for C_{10} has a worst-case error of 2% and an average error of 0.4% over this range; the corresponding numbers for L are 7% and 1.7%. The uncertainty in C_{10} is approximately 15% for all models. The likelihood reaches its maximum at the point $(D_1', D_1'') = (0.0, -3.0)$, which is beyond the range covered by the fitting formula. The peak likelihood for this model is 3.7 times the likelihood of a flat Harrison-Zel'dovich model. Figure 5.4 shows L as a function of D_1' and D_1''.

In Tables 5.2 and 5.3 we show the best-fitting shape parameters for some flat, low-Ω_0 variants of the CDM model. The angular power spectra that were used in fitting to the shape parameters were computed by Martin White, as were the small-scale normalizations σ_8 and Δ^2. The fit of a quadratic to these theories gives an error at the worst-fit multipole (in the range $2 \leq l \leq 30$) of about 5%, with a typical error of $\lesssim 2\%$, showing that such theories are fitted well by quadratics over the range of scales probed by COBE. To quantify the error introduced by approximating the power spectrum as a quadratic, we computed likelihood curves for the *worst*-fitting model in our sample using both the true power spectrum and the quadratic approximation. The two curves differ by 11% in peak likelihood and by 0.5% in normalization.

Critical ($\Omega_0 = 1$) CDM models with late reionization are also well fit by quadratic $D(x)$. The most plausible, though not the only, ionization history in hierarchical models of structure formation is standard recombination, followed by full ionization from some redshift z_* until the present. The fully ionized phase is due (perhaps) to radiation from massive stars on scales that go non-linear early [113]. We find that, over the range of l probed by COBE, models that have $z_* \lesssim 100$ are almost indistinguishable from models with no reionization, assuming standard big-bang nucleosynthesis values for Ω_B. There is, of course, damping on degree scales ($l \sim 100$), but little change in the spectrum at smaller

Ω_0	n	D_1'	D_1''	Δ^2	σ_8	D_1'	D_1''	Δ^2	σ_8
0.50	0.85	-0.079	0.476	0.027	1.45	-0.186	0.910	0.015	1.08
	0.90	0.046	0.503	0.031	1.62	-0.043	0.785	0.020	1.31
	0.95	0.172	0.557	0.036	1.82	0.115	0.693	0.028	1.62
	1.00	0.298	0.635	0.040	2.05	0.298	0.635	0.040	2.05
0.40	0.85	-0.110	0.600	0.030	1.29	-0.202	0.955	0.016	0.95
	0.90	0.012	0.631	0.034	1.45	-0.065	0.853	0.022	1.15
	0.95	0.135	0.685	0.039	1.62	0.084	0.787	0.030	1.43
	1.00	0.258	0.767	0.045	1.82	0.258	0.767	0.045	1.82
0.30	0.85	-0.159	0.853	0.032	1.10	-0.225	1.058	0.016	0.79
	0.90	-0.040	0.885	0.036	1.24	-0.097	0.991	0.022	0.97
	0.95	0.080	0.944	0.041	1.39	0.041	0.975	0.031	1.21
	1.00	0.201	1.022	0.047	1.56	0.201	1.022	0.047	1.56
0.20	0.85	-0.240	1.338	0.029	0.82	-0.258	1.237	0.014	0.56
	0.90	-0.124	1.366	0.033	0.92	-0.143	1.229	0.019	0.69
	0.95	-0.008	1.415	0.038	1.03	-0.023	1.295	0.027	0.88
	1.00	0.108	1.485	0.043	1.15	0.108	1.485	0.043	1.15
0.10	0.85	-0.342	2.455	0.014	0.33	-0.275	1.561	0.006	0.21
	0.90	-0.227	2.462	0.016	0.37	-0.176	1.673	0.008	0.26
	0.95	-0.113	2.485	0.019	0.41	-0.082	1.931	0.012	0.33
	1.00	0.000	2.527	0.022	0.46	0.000	2.527	0.022	0.46

Table 5.2: Shape parameters for ΛCDM models

We give the shape parameters D_1' and D_1'' of the radiation power spectrum in a ΛCDM model with $h = 0.75$. Also shown is the matter power spectrum normalization with the radiation normalized to $C_{10} = 10^{-11}$. For the tilted models we show the results with (right columns) and without (left columns) a gravity-wave component with $C_2^T/C_2^S = 7(1-n)$.

Ω_0	n	D_1'	D_1''	Δ^2	σ_8	D_1'	D_1''	Δ^2	σ_8
0.50	0.85	-0.001	0.657	0.018	0.80	-0.123	1.039	0.010	0.60
	0.90	0.130	0.698	0.020	0.89	0.028	0.937	0.013	0.72
	0.95	0.262	0.769	0.023	1.00	0.196	0.877	0.018	0.89
	1.00	0.396	0.866	0.026	1.12	0.396	0.866	0.026	1.12
0.40	0.85	-0.018	0.807	0.018	0.67	-0.130	1.099	0.010	0.50
	0.90	0.111	0.851	0.021	0.75	0.016	1.021	0.013	0.60
	0.95	0.240	0.924	0.024	0.84	0.178	0.992	0.019	0.74
	1.00	0.372	1.025	0.027	0.94	0.372	1.025	0.027	0.94
0.30	0.85	-0.043	1.104	0.017	0.51	-0.140	1.222	0.009	0.36
	0.90	0.084	1.154	0.020	0.56	-0.000	1.188	0.012	0.44
	0.95	0.212	1.227	0.022	0.63	0.155	1.212	0.017	0.55
	1.00	0.340	1.327	0.026	0.70	0.340	1.327	0.026	0.70
0.20	0.85	-0.071	1.668	0.013	0.32	-0.147	1.436	0.006	0.22
	0.90	0.055	1.719	0.015	0.35	-0.015	1.471	0.008	0.27
	0.95	0.180	1.788	0.017	0.39	0.131	1.594	0.012	0.34
	1.00	0.307	1.883	0.019	0.44	0.307	1.883	0.019	0.44
0.10	0.85	-0.029	2.910	0.003	0.13	-0.112	1.806	0.001	0.08
	0.90	0.096	2.946	0.004	0.14	0.018	1.978	0.002	0.10
	0.95	0.222	2.998	0.004	0.15	0.163	2.326	0.003	0.13
	1.00	0.347	3.072	0.005	0.17	0.347	3.072	0.005	0.17

Table 5.3: More shape parameters for ΛCDM models

As in Table 1 but with $h = 0.50$.

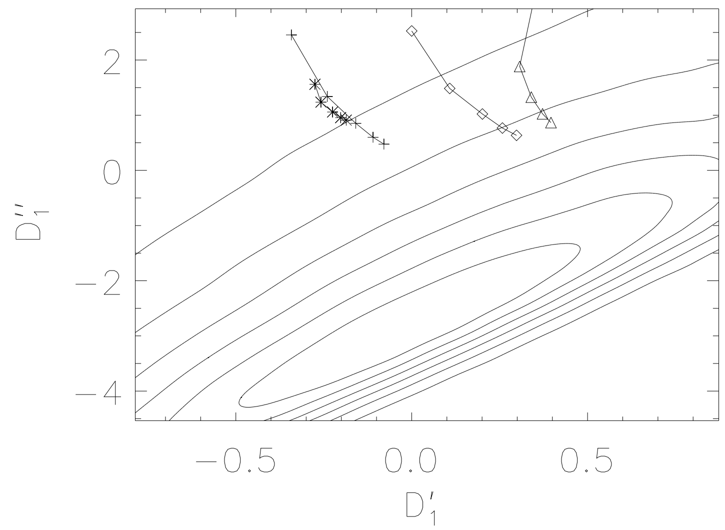

Figure 5.4: Likelihoods for quadratic models

The likelihood, L, is plotted as a function of the power spectrum shape parameters D_1' and D_1''. The contours range from $L = 0.5$ to $L = 3$ in steps of 0.5, where $L = 1$ corresponds to a flat spectrum $D_1' = D_1'' = 0$. For values of (D_1', D_1'') for which the power spectrum goes negative over the range $2 \leq l \leq 30$ (lower right corner) we have set the likelihood to zero. We have also indicated the values of (D_1', D_1'') for several of the models in Tables 5.2 and 5.3. The crosses and asterisks are for $n = 0.85$ models with and without gravity waves, respectively. Both of these models have $h = 0.75$. The diamonds and triangles denote $n = 1$ models with $h = 0.75$ and $h = 0.5$, respectively. In all cases, Ω ranges from 0.1 to 0.5 in steps of 0.1, and D_1'' decreases with increasing Ω.

l. Further, the relative normalization of the matter and radiation power spectra is the same as in models with standard recombination. For a CDM model with $h = 0.5$ and $n = 1$, the quadratic parameters are fitted well by the formulae

$$D_1' = 0.738 - 0.0307 z_* + 3.32 \times 10^{-4} z_*^2 - 1.06 \times 10^{-6} z_*^3, \qquad \text{and} \tag{5.13}$$

$$D_1'' = 1.554 - 0.0483 z_* + 3.67 \times 10^{-4} z_*^2 - 7.60 \times 10^{-7} z_*^3, \tag{5.14}$$

over the range $30 \leq z_* \leq 110$.

The case of open CDM models is more complicated, since the C_l exhibit features at several scales. To add to the difficulty, there appear to be several different primordial spectra one can consider in open universe models. Some models based on inflationary phases [121, 144, 29] predict power spectra that show an increase in power near the curvature radius. All of these calculations use basis functions in which there is exponential damping of power above the curvature radius; however, this assumption can be relaxed [122, 197]. The open models of Ratra & Peebles [144] have already been fitted to the 2-year COBE data [72], and we will not duplicate the results here. We mention, however, that the C_l for such models can be fitted by cubics in $\log_{10} l$ to the same accuracy that the ΛCDM models can be fitted by quadratics. This increases the dimension of the parameter space and makes tabulating the results more difficult. We defer consideration of cubic fits until the situation with regard to open models is more settled.

Once the 4-year COBE data become available, we hope that fitting formulae similar to equations (5.11) and (5.12), but going to sufficiently high order to encompass almost all theories, could be produced for the benefit of the astrophysics community. Such a fit, coded into a subroutine, would allow any theory to be quickly and accurately fitted to the COBE data. At present our simple quadratic fit is sufficient for a wide range of theories of current interest.

The best normalization and the goodness of fit of the temperature fluctuations for a range of models are given by equations (5.11) and (5.12). Using these results to normalize the matter power spectrum from the CMB can present some complications. In the simplest picture, in which the large-angle CMB anisotropy comes purely from potential fluctuations on the last scattering surface, the relative normalization of the CMB and matter power spectrum today is straightforward, as we saw in the previous section. For models with $\Omega_0 < 1$ the normalization is not so straightforward. Naively one would think that, for fixed CMB fluctuations at $z = 1000$, one would have smaller matter fluctuations today. This is because in an open or a flat model with a cosmological constant ($\Omega_\Lambda = 1 - \Omega_0$), density perturbations stop growing once the universe becomes either curvature or cosmological constant dominated (respectively). Curvature domination occurs quite early, and the growth of density fluctuations $\delta\rho/\rho \equiv \delta$ in an open universe is suppressed (relative to an $\Omega_0 = 1$ universe) by a factor $\Omega_0^{0.6}$. In a flat Λ model, the cosmological constant dominates only at late times, and so the growth-suppression is a weaker function of the matter content: $\delta \propto \Omega_0^{0.23}$ [36]. This suppression of growth in an $\Omega_0 < 1$ universe has often been cited as "evidence" that Ω_0 must be large — otherwise fluctuations could not have grown enough to form the structures we observe today.

In fact there are several other effects that come into play when normalizing the matter power spectrum to the COBE data in a low-Ω_0 model. The first is that, though the growth in such models is suppressed by Ω_0^p ($p \simeq 0.6$ for open and 0.23 for Λ models (for a more general formula, see [36]), the potential fluctuations are proportional to Ω_0. Hence the CMB fluctuations are even more suppressed than are the density fluctuations! So for a fixed COBE normalization the matter fluctuations today are *larger* in a low-Ω_0 universe, and the cosmological constant model clearly has the most enhancement, since the fluctuation growth is the least suppressed. In terms of the power spectrum, $P(k)$, we expect for fixed COBE normalization that $P(k) \propto \delta^2 \propto \Omega_0^{2(p-1)}$, as has been pointed out in [59].

This potential suppression is not the only effect that occurs in low-Ω_0 universes, although it is the largest. Due to the fact that the fluctuations stop growing (or, in other words, the potentials decay) at some epoch, there is another contribution to the large-angle CMB anisotropy measured by COBE. In addition to the redshift experienced while climbing out of potential wells on the last scattering surface, photons experience a cumulative energy change due to the decaying potentials as they travel to the observer. If the potentials are decaying, the blueshift of a photon falling into a potential well is not entirely canceled by a redshift when it climbs out. This leads to a net energy change, which accumulates along the photon path. This is often called the Integrated Sachs-Wolfe (ISW) effect, to distinguish it from the more commonly considered redshifting that has become known as the Sachs-Wolfe effect, although both effects were considered in the paper of Sachs & Wolfe [149]. This ISW effect will operate most strongly on scales where the change of the potential is large over a wavelength, so preferentially on large angles [104].

In Λ models, the ISW effect can change the relative normalization of the matter and radiation fluctuations at the 25% level for $\Omega_0 \sim 0.3$. We show in Figure 5.5 how these various effects on the inferred matter power spectrum normalization scale with Ω_0 in a cosmological constant universe (the simplest case). We see the total power is slightly changed, for fixed C_{10}, because the shape of the C_l depend on Ω_0. This affects the goodness of fit to the COBE data [32], as we shall discuss in more detail in the next section. The ratio of the large-scale matter normalization to C_{10} is changed by the ISW contribution to C_{10}, the change in the potentials and the growth of fluctuations from $z = 1000$ to the present. Over the range $\Omega_0 = 0.1$ to 0.5 one finds for an $n = 1$ spectrum with $C_{10} = 10^{-11}$ that

$$\lim_{k\to 0} \frac{P(k)}{k} = 1.14 \times 10^6 \Omega_0^{-1.35} \quad (h^{-1}\mathrm{Mpc})^4, \tag{5.15}$$

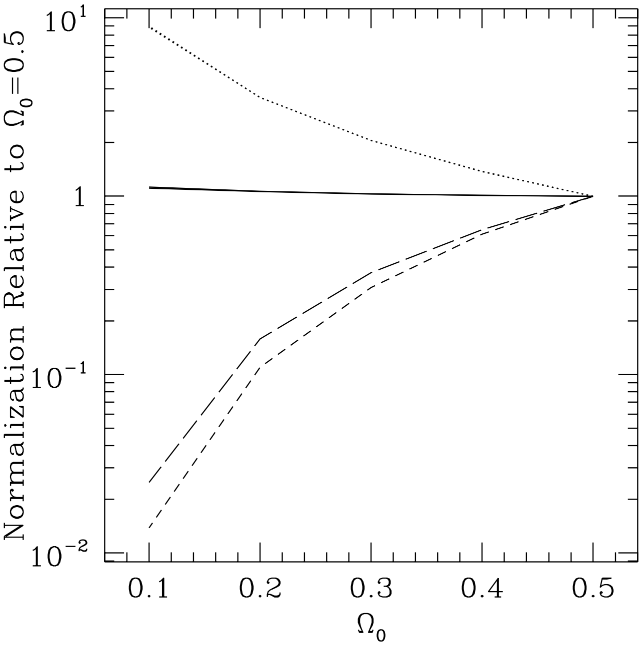

Figure 5.5: Contributions to the normalization of ΛCDM models

All quantities are normalized to their values at $\Omega_0 = 0.5$. The solid line is the RMS temperature fluctuation with C_{10} fixed. The dotted line shows the ratio of the large-scale matter normalization ($\lim_{k\to 0} P(k)/k$) to C_{10}. The dashed lines show the effect of the shift in matter-radiation equality on the small-scale normalization σ_8, holding the large-scale normalization $\lim_{k\to 0} P(k)/k$ fixed. We show two models: $h = 0.75$ (upper) and $h = 0.50$ (lower) with $\Omega_B h^2 = 0.0125$.

almost independent of h. This can be compared with the scaling presented above. Also, the epoch of matter-radiation equality is shifted, which changes the normalization on smaller scales for fixed large-scale $P(k)$. Putting these effects together, we show the RMS fluctuation on a scale $0.028\,h\mathrm{Mpc}^{-1}$ (see below) as a function of Ω_0 in Figure 5.6. The sharp downturn at low Ω_0 is due to a combination of the larger scale of matter-radiation equality, moving the break in the power spectrum to smaller k, and the photon drag on the baryons having an increased effect on fluctuation growth for large Ω_B/Ω_0. For $\Omega_0 \simeq 0.3$ the shift in matter-radiation equality and the scaling of equation (5.15) roughly cancel, making Δ^2 much less sensitive to Ω_0 than the individual contributions would suggest.

We note here that the shape of the C_l for the tensor (gravitational wave) modes is

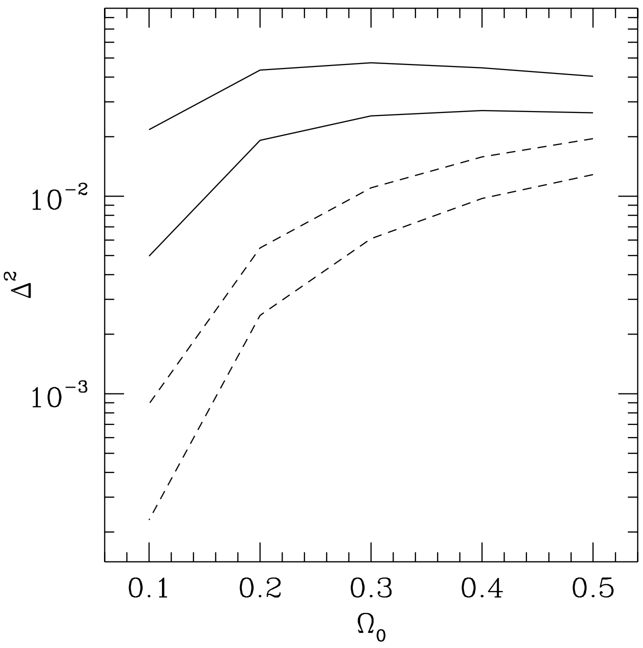

Figure 5.6: Matter fluctuations in low-density CDM models

The normalization, $\Delta^2(0.028h\mathrm{Mpc}^{-1})$, is plotted as a function of Ω_0 for open CDM (dashed) and ΛCDM (solid) models normalized to $C_{10} = 10^{-11}$. In both cases the upper curves are for $h = 0.75$ and the lower curves are for $h = 0.50$, both with $\Omega_B h^2 = 0.0125$.

largely independent of Ω_0 [178]. For this reason the radiation power spectrum of a Λ model with some tilt and a component of tensors can exhibit less curvature at $l \sim 10$ than a purely scalar power spectrum. Since in some inflationary models we expect a non-negligible tensor component [48], we have computed the tensor C_l following Crittenden *et al.* [44] and give results both including and excluding a significant tensor contribution. Our results update those of Kofman, Gnedin & Bahcall [103], who also considered tilted, ΛCDM models with a component of gravity waves.

In Tables 5.2 and 5.3, and in Figure 5.6, we show the normalization of the matter power spectrum for a range of models, holding the CMB normalization (C_{10}) fixed. We quote both the value of the RMS density fluctuation at $0.028h\mathrm{Mpc}^{-1}$ (large-scale) and σ_8

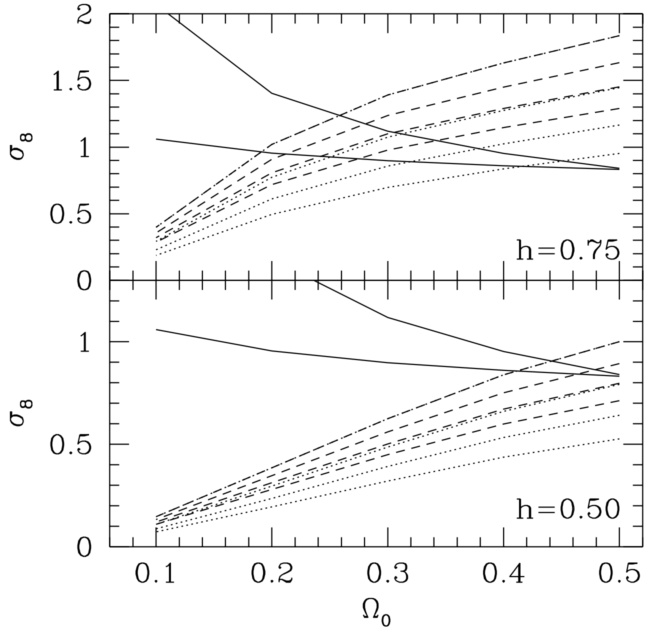

Figure 5.7: σ_8 in ΛCDM models

The dashed lines are for models in which the entire contribution to the temperature anisotropy measured by COBE comes from scalar perturbations; the dotted lines are for models with both scalars and tensors, with $C_2^T/C_2^S = 7(1-n)$. The slope n increases from 0.85 to 1.00 in steps of 0.05 with the lowest n yielding the lowest σ_8. The two solid lines are two observational determinations of σ_8, the top line from cluster abundances and the bottom line from large-scale structure [equations (5.17) and (5.18)].

(small-scale). For comparison, Peacock & Dodds give [133]

$$\Delta^2\left(k = 0.028h\mathrm{Mpc}^{-1}\right) \equiv \frac{d\sigma_\rho^2}{d\ln k} = (0.0087 \pm 0.0023)\,\Omega_0^{-0.3}. \tag{5.16}$$

There are many determinations of σ_8; here we quote those from Peacock & Dodds [133]

$$\sigma_8 = 0.75\Omega_0^{-0.15} \tag{5.17}$$

and cluster abundances [191]

$$\sigma_8 = 0.57\Omega_0^{-0.56} \tag{5.18}$$

where the scaling with Ω_0 in both cases is for models with $\Omega_0 + \Omega_\Lambda = 1$. These values are consistent with those inferred from large-scale flows [49] and direct observations [120].

In Figure 5.7 we compare these observations with the COBE-normalized values of σ_8 for a range of ΛCDM models.

5.4 Constraining the Cosmological Constant

It has been known for some time that standard CDM models predict too much small-scale power when normalized to COBE at large scales [59, 133]. In order to resolve this problem, several modifications have been proposed to change slightly the shape of the power spectrum. Among the proposed fixes are tilting the primordial spectrum [48], mixing some hot dark matter in with the cold [47], and assuming a very small value for the Hubble constant [9]. As we have seen in the previous section, one of the most promising modifications involves replacing some of the cold dark matter with a cosmological constant Λ while keeping the geometry flat (see, *e.g.*, [59, 60, 103]). This has the added advantage of solving the cosmic age problem.

It is known, however, that Λ models produce quite different CMB anisotropies on large scales [104] because of the integrated Sachs-Wolfe effect [149]. Sugiyama & Silk [171] numerically solved the perturbation equations, and obtained CMB power spectra for different Λ's. The power spectra for Λ-dominated models have shapes that are quite different from the simple power-law models that are usually used in analyzing the CMB anisotropy. Therefore, analyses based on the simple Sachs-Wolfe power-law spectrum (2.29) (*e.g.*, [155, 163, 71]) cannot be used to set limits on Λ. A direct comparison between the DMR data and the predicted power spectra from the Λ models is required.

In this section we compare predicted CMB power spectrum of ΛCDM models with the COBE DMR data. In these models, the Universe is supposed to be cosmologically flat, so that the density parameter Ω and the cosmological constant Λ add up to unity. The matter in the Universe is assumed to consist primarily of CDM, with a few percent of the matter being baryonic. It should be noticed that the CMB anisotropy on the angular scales probed by COBE is largely independent of the matter content of the Universe. Therefore our results obtained here are general for ΛCDM models. The initial fluctuations are supposed to be Gaussian and adiabatic, with a Harrison-Zel'dovich ($n = 1$) power spectrum, as in standard CDM. These models fit the galaxy clustering data quite well with $\Lambda \sim 0.8$ [59, 60, 103].

We have applied the Karhunen-Loève technique presented in the previous chapter to the COBE DMR two-year sky maps. Figure 5.8 shows the angular power spectra C_l for

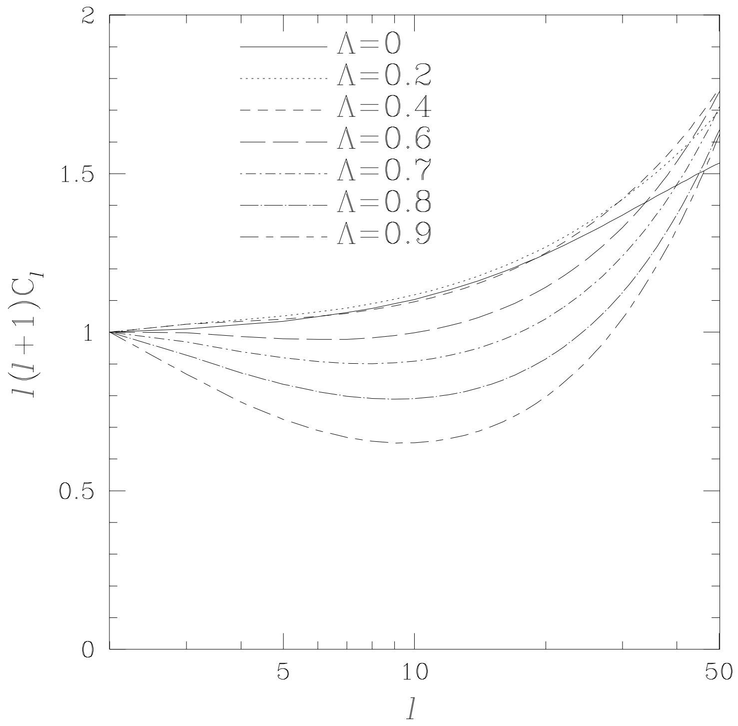

Figure 5.8: Angular power spectra for ΛCDM models

The angular power spectrum $l(l+1)C_l$ is shown for each of the seven ΛCDM models described in the text. The power spectra have been normalized so that $6C_2 = 1$.

seven ΛCDM models. These power spectra were obtained by direct numerical calculations up to the present epoch, using a Boltzmann code by Naoshi Sugiyama (see [170]). All have Hubble constant $H_0 = 50\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and baryon density $\Omega_B = 0.03$. The cosmological constant Λ takes the values $0, 0.2, 0.4, 0.6, 0.7, 0.8, 0.9$. (The model with $\Lambda = 0$ is, of course, standard CDM.) The dependence on Ω_B of C_l is quite weak, i.e., less than 10% for $l \lesssim 30$. The dependence on the Hubble constant is also weak: changing H_0 from 50 to 80 leaves the C_l's almost identical for $l \lesssim 10$, and causes less than a 20% difference for $l \lesssim 30$. It is clear from Figure 5.8 that the C_l's for these models are quite different from flat or simple power-law models, particularly when Λ is large. For each of these models, we computed L as a function of the power spectrum normalization $Q_{\rm ps}$. These likelihoods are shown in Figure 5.9. It is clear that the data prefer low values of Λ.

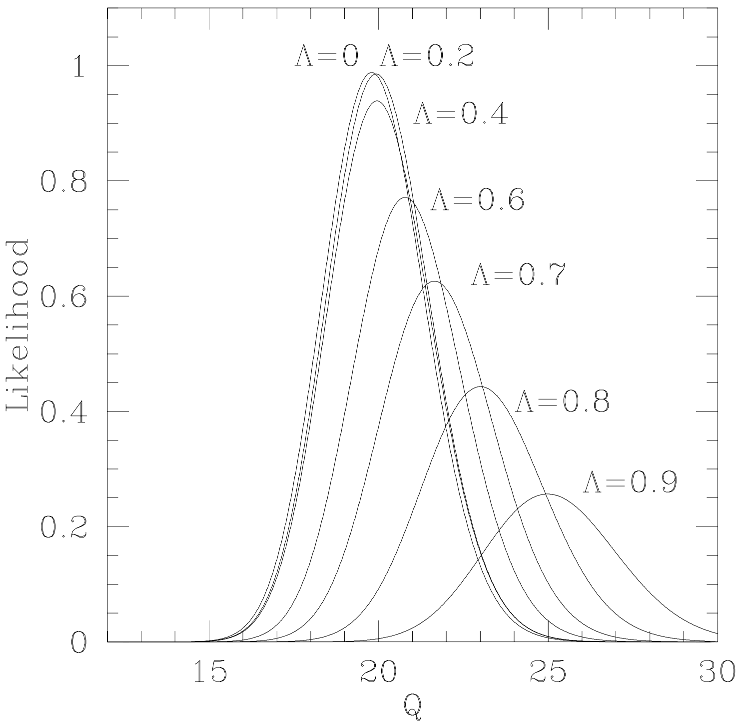

Figure 5.9: Likelihood functions for ΛCDM models

The likelihood L is plotted as a function of Q_{ps} for each of the seven ΛCDM models described in the text. The overall normalization is arbitrary.

We can place Bayesian confidence limits on Λ in the following way. For each of the seven models, we compute the marginal likelihood by integrating over the normalization Q_{ps}:

$$L_{\text{marg}}(\Lambda) \equiv \int L(\Lambda, Q_{\text{ps}})\, dQ_{\text{ps}}. \tag{5.19}$$

These marginal likelihoods are plotted in Figure 5.10 as a function of Λ. We then smoothly interpolate between the data points, using a natural cubic spline. We then say that Λ has an upper bound Λ_{max} at some confidence level c if

$$\int_0^{\Lambda_{\text{max}}} L_{\text{marg}}(\Lambda)\, d\Lambda = c \int_0^1 L_{\text{marg}}(\Lambda)\, d\Lambda. \tag{5.20}$$

The upper limit on Λ derived in this way is the boundary of a Bayesian credible region in the Q_{ps}-Λ plane, adopting a prior distribution that is uniform in both Q_{ps} and Λ [15]. A uniform distribution is a natural choice of prior for Λ. Several different choices of prior distribution

Confidence level	Λ_{max} including Q	Λ_{max} excluding Q
68%	0.55	0.62
90%	0.78	0.85
95%	0.86	0.92
99%	0.96	0.98

Table 5.4: Upper limits on Λ

could be justified for $Q_{\rm ps}$; however, since the data contain a very strong detection, the results are insensitive to the prior in Q [33]. Table 5.4 shows the upper limits on Λ for several different confidence levels.

An alternative procedure would be to use the maximum value of the likelihood, $L_{\rm max}(\Lambda) \equiv {\rm max}_Q L(\Lambda, Q)$, in place of the marginal likelihood in equation (5.20). In practice, it makes very little difference which procedure is followed: for example, replacing $L_{\rm marg}$ by $L_{\rm max}$ changes the 95% confidence level upper limit on Λ from 0.86 to 0.87.

The quadrupole moment of the DMR sky maps is anomalously low, and it is has been suggested that it may be contaminated in some way (see *e.g.*, [71]). In Section 2.7 we argued against the presence of contamination, and, in particular, against the idea that the quadrupole information should be thrown away entirely. Nonetheless, in order to conform to standard practice and in order to assess the degree of sensitivity of our results to the possibility of contamination, we removed the quadrupole from the data, and repeated the analysis including only modes with $l \geq 3$, marginalizing over the quadrupole as well as the monopole and dipole. Table 5.4 contains the upper limits on Λ found in this way. Since the large quadrupole predicted by ΛCDM models is at odds with the low quadrupole in the data, it is not surprising that removing the quadrupole weakens the constraints. However, as we argue in Section 2.7, we should be wary of excising the quadrupole from the data simply because it is anomalously low: such a biased editing procedure is clearly at odds with sound statistical practice.

The likelihood curve in Figure 5.10 is flat near $\Lambda = 0$. The reason for this is easily seen by looking at Figure 5.8: the angular power spectrum for $\Lambda = 0.2$ is almost identical to the $\Lambda = 0$ model, indicating that C_l is a very weak function of Λ for small Λ. Specifically, if we perform a Taylor expansion of C_l about $\Lambda = 0$, the linear term is negligibly small.

In conclusion, we find that ΛCDM models are strongly disfavored by the DMR data.

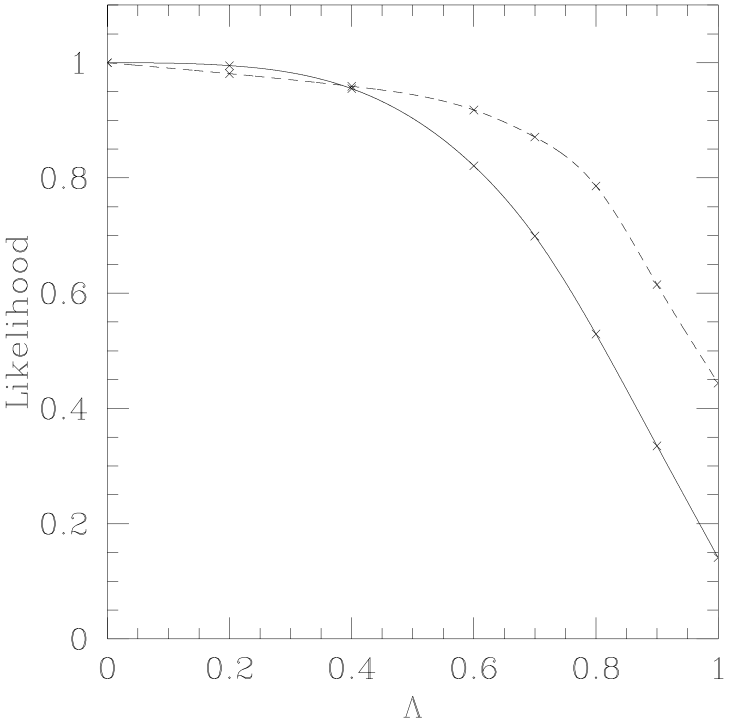

Figure 5.10: Marginal likelihoods for ΛCDM models

The marginal likelihood $L_{\rm marg}$ is plotted as a function of Λ. The solid curve is the likelihood including the quadrupole, and the dashed curve is the result of excluding the quadrupole. The crosses show the marginal likelihood for each of the seven models described in the text. The smooth curves are cubic spline interpolations between them. The overall normalization is arbitrary.

In particular, models with $\Lambda > 0.86$ are inconsistent with the data at a confidence level of 95%, while models with $\Lambda > 0.78$ are inconsistent at the 90% level. The large-scale structure data seem to prefer ΛCDM models with $\Lambda \gtrsim 0.7$. It is clear from these results that the DMR data are at best marginally consistent with such models.

Removing the quadrupole from the data weakens these constraints somewhat. However, it is important to note that the primary reason for removing the quadrupole is simply that it is anomalously low. Removing data points simply because they do not fit our theoretical expectations is a dangerous statistical practice. In the absence of strong independent evidence that the quadrupole is contaminated, one should be wary of throwing it away.

These results we obtained here are about the same as the best current limits from gravitational lensing [101, 102]. It should be noticed that our method is totally independent and free from any ambiguities involved in the gravitational lensing method, such as the evolution and mass distribution of galaxies, and therefore provides a valuable confirmation of it. There is every reason to expect that, once the full four-year DMR data have been analyzed, it will be possible to discriminate between models with $\Lambda \approx 0.6$ and those with $\Lambda = 0$ with very high confidence, since the power spectra in these models have very different shapes (see Figure 5.8).

Chapter 6

Wiener Filtering of the COBE Data

> Raw data, like raw sewage, requires some processing before it can be spread around.
>
> James A. Carr

6.1 Introduction

Until now, our analysis of the COBE DMR data has focused primarily on the problem of estimating the power spectrum. In fact, most workers in the field have regarded this as the primary use for the DMR data [13, 20, 21, 31, 32, 50, 71, 151, 155, 174, 185, 194, 195, 196], although a few have focused on other problems, such as testing the hypothesis that the anisotropy obeys Gaussian statistics [82, 107, 164, 177]. In addition, statistical comparisons have been made between the DMR sky maps and data from other experiments [66, 183]. Less attention has been paid to studying the spatial properties of the sky maps themselves to determine, for example, which hot and cold spots on the maps are likely to be real, and which are dominated by noise. Such an analysis has the potential to be quite useful in comparing the DMR results to other experiments that probe comparable angular scales, such as the FIRS [125] and Tenerife [45, 79, 182] experiments. In this chapter, we will use Wiener filtering to study these questions, extending a similar study that we did for the first-year DMR maps [34].

There are two major difficulties in studying the DMR maps. The signal-to-noise ratio in each pixel is quite low, and pixels near the Galactic plane are contaminated by Galactic emission. These problems place strict limits on the accuracy with which the true structure can be recovered [30, 167]. In this chapter we apply the techniques of Wiener filtering and "constrained realizations" in an attempt to mitigate these problems. We assume the correctness of the canonical theory of large-scale CMB anisotropy: that the anisotropy forms a Gaussian random field with a known power spectrum. We then remove the pixels that are presumed to be contaminated, and apply an optimal linear filter to the data in an attempt to clean up the noise and see the underlying structure. The technique of constrained realizations helps quantify the uncertainties associated with this method.

The Wiener filtering described in this chapter is based on ideas developed in a more general context by Rybicki and Press [148], and it is similar to a technique applied recently to galaxy catalogues [110, 198]. The formalism for making constrained realizations of Gaussian random fields has also been discussed elsewhere [67, 83], as has the connection between this formalism and the Wiener filter [198].

The idea behind the Wiener filter is quite simple. If we are willing to assume that we know the statistical properties of the signal and noise contributions to the data, we can apply a linear filter that preferentially removes noise and leaves signal. The Wiener filter is the optimal linear filter for this purpose. It can be derived in two independent ways:

1. It is the optimal linear filter in the sense of least squares. That is, the Wiener filter minimizes the ensemble average of the mean-square deviation between the filtered data and the underlying cosmic signal.
2. It is a maximum-likelihood estimator of the cosmic signal. That is, the filtered sky map is the most likely realization of the underlying signal at each point.

As long as both the signal and the noise are Gaussian, the filters chosen by these two criteria are identical. Furthermore, we can determine the complete *a posteriori* probability distribution of the true signal given the data and the assumed power spectrum. This probability distribution is Gaussian, with mean given by the filtered data. We can compute the covariance matrix of this probability distribution and use it to construct "constrained realizations" of the CMB anisotropy with the correct *a posteriori* probability distribution. These realizations are useful for qualitatively assessing the uncertainties in the filtered maps, and for using Monte Carlo simulations to make predictions for other experiments.

We will derive all of these results in the next two sections. We will then apply this formalism to the two-year DMR sky maps to identify features in the data that have high statistical significance. Finally, we will use the filtered data to make predictions for the Tenerife experiment, which probes slightly smaller angular scales than the DMR.

6.2 Two Derivations of the Wiener Filter

As in Section 2.6, we will represent our data set as an $\mathcal{N}$-dimensional vector $\vec{d}$, containing contributions from both signal and noise:

$$\vec{d} = \vec{s} + \vec{n} = \mathbf{Y} \cdot \mathbf{W} \cdot \vec{a} + \vec{n}. \tag{6.1}$$

(For the moment, we ignore complications having to do with the unknown monopole and dipole contributions; we will see how to account for them properly below.) a_μ is a coefficient of the spherical harmonic expansion of the true signal $\vec{s}$. (Here we are using Greek indices to represent pairs of spherical harmonic indices (lm) as usual.) $Y_{i\mu} = Y_\mu(\hat{\mathbf{r}}_i)$ is a spherical harmonic evaluated at the position of the ith pixel, and n_i is the noise in the ith pixel. $\mathbf{W}$ is a diagonal matrix whose entries are those of the experimental window function described in Section 2.4. We make the usual assumption of Gaussian statistics: both a_μ and n_i are Gaussian random variables with zero mean and covariances

$$\langle \vec{a} \cdot \vec{a}^T \rangle = \mathbf{C}, \tag{6.2}$$

$$\langle \vec{n} \cdot \vec{n}^T \rangle = \mathbf{N}, \tag{6.3}$$

$$\langle a_\mu n_i \rangle = 0. \tag{6.4}$$

The matrices $\mathbf{C}$ and $\mathbf{N}$ are diagonal, with $C_{\mu\mu} = C_l$ and $N_{ii} = \sigma_i^2$. In principle, $\vec{a}$ and $\mathbf{C}$ should be infinite-dimensional, but in practice the COBE beam smooths away modes with high l, and we can truncate the spherical harmonic expansion at some $l_{\max}$. We shall see below that the filter recovers virtually no information for high l, and the results are therefore quite insensitive to $l_{\max}$. From Figure 6.7, below, we can see that it is reasonable to set $l_{\max} = 30$, which means that the dimension of $\vec{a}$ is $\mathcal{M} = (l_{\max} + 1)^2 = 961$. The covariance matrix of the data vector $\vec{d}$ contains contributions from both signal and noise:

$$\mathbf{M} \equiv \langle \vec{d} \cdot \vec{d}^T \rangle = \mathbf{Y} \cdot \bar{\mathbf{C}} \cdot \mathbf{Y}^T + \mathbf{N}, \tag{6.5}$$

where $\bar{\mathbf{C}} = \mathbf{W} \cdot \mathbf{C} \cdot \mathbf{W}$.

We will now derive the Wiener filter as a least-squares estimator. We wish to find a linear filter $\mathbf{F}$ that will enable us to estimate the true cosmic signal $\vec{s}$ as closely as possible. Specifically, if we set

$$\vec{y} = \mathbf{F} \cdot \vec{d}, \tag{6.6}$$

then we want to choose the $\mathcal{N} \times \mathcal{N}$ matrix $\mathbf{F}$ to minimize

$$\Delta \equiv \langle (\vec{y} - \mathbf{Y} \cdot \vec{a})^2 \rangle. \tag{6.7}$$

Combining these two equations, we find that

$$\Delta = \mathrm{Tr}\left[(\mathbf{F} \cdot \mathbf{Y} \cdot \mathbf{W} - \mathbf{Y}) \cdot \mathbf{C} \cdot (\mathbf{F} \cdot \mathbf{Y} \cdot \mathbf{W} - \mathbf{Y})^T + \mathbf{F} \cdot \mathbf{N} \cdot \mathbf{F}^T \right]. \tag{6.8}$$

Setting $\partial \Delta / \partial F_{ij} = 0$ and solving for $\mathbf{F}$, we find the least-squares filter

$$\mathbf{F} = \mathbf{Y} \cdot \mathbf{C} \cdot \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{M}^{-1}. \tag{6.9}$$

This derivation makes no use of the fact that the random variables $\vec{a}$ and $\vec{n}$ are Gaussian; it uses only their covariances. If we make use of our assumption that these quantities are Gaussian, we can give an alternative characterization of the filter (6.9): it gives the maximum-likelihood estimator of the signal $\vec{s}$.

Bayes's theorem states that the conditional probability density of $\vec{a}$ being the correct set of spherical harmonic coefficients given the data $\vec{d}$ is simply

$$f(\vec{a} \,|\, \vec{d}) \propto f(\vec{d} \,|\, \vec{a}) f(\vec{a}), \tag{6.10}$$

where $f(\vec{d} \,|\, \vec{a})$ is the probability density that the actual data $\vec{d}$ would have occurred assuming $\vec{a}$ is the correct set of coefficients, and $f(\vec{a})$ is the *a priori* probability of getting the coefficients $\vec{a}$. All of these probability densities are Gaussian:

$$f(\vec{a} \,|\, \vec{d}) \propto \exp\left(-\tfrac{1}{2} (\vec{d} - \mathbf{Y} \cdot \mathbf{W} \cdot \vec{a})^T \cdot \mathbf{N}^{-1} \cdot (\vec{d} - \mathbf{Y} \cdot \mathbf{W} \cdot \vec{a}) \right) \exp\left(-\tfrac{1}{2} \vec{a}^T \cdot \mathbf{C}^{-1} \cdot \vec{a} \right). \tag{6.11}$$

By completing the square and performing some unenlightening algebraic manipulations, one can show that $f(\vec{a} \,|\, \vec{d}) \propto \exp\left(-\frac{1}{2} \chi^2 \right)$, where

$$\chi^2 = (\vec{a} - \vec{a}_{\mathrm{ML}})^T \cdot (\mathbf{C}^{-1} + \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{Y} \cdot \mathbf{W}) \cdot (\vec{a} - \vec{a}_{\mathrm{ML}}) + \vec{d}^T \cdot \mathbf{M}^{-1} \cdot \vec{d}, \tag{6.12}$$

and

$$\vec{a}_{\mathrm{ML}} = (\mathbf{C}^{-1} + \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{Y} \cdot \mathbf{W})^{-1} \cdot \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \vec{d}. \tag{6.13}$$

It is easy to see from equation (6.12) that the maximum-likelihood value of $\vec{a}$ is $\vec{a}_{\rm ML}$. The maximum-likelihood estimate of the actual CMB signal is therefore $\vec{s}_{\rm ML} = \mathbf{Y} \cdot \vec{a}_{\rm ML} = \mathbf{F}' \cdot \vec{d}$, where the filtering matrix is

$$\mathbf{F}' = \mathbf{Y} \cdot \left(\mathbf{C}^{-1} + \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{Y} \cdot \mathbf{W}\right)^{-1} \cdot \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1}. \tag{6.14}$$

Despite their different appearances, in fact $\mathbf{F} = \mathbf{F}'$:

$$\begin{aligned}
\mathbf{F}' &= \mathbf{Y} \cdot \left(\mathbf{C}^{-1} + \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{Y} \cdot \mathbf{W}\right)^{-1} \cdot \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{M} \cdot \mathbf{M}^{-1} \\
&= \mathbf{Y} \cdot \left(\mathbf{C}^{-1} + \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{Y} \cdot \mathbf{W}\right)^{-1} \cdot \\
&\qquad \left(\mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{Y} \cdot \mathbf{W} \cdot \mathbf{C} \cdot \mathbf{W} \cdot \mathbf{Y}^T + \mathbf{W} \cdot \mathbf{Y}^T\right) \cdot \mathbf{M}^{-1} \\
&= \mathbf{Y} \cdot \left(\mathbf{C}^{-1} + \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{Y} \cdot \mathbf{W}\right)^{-1} \cdot \\
&\qquad \left(\mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{Y} \cdot \mathbf{W} + \mathbf{C}^{-1}\right) \cdot \mathbf{C} \cdot \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{M}^{-1} \\
&= \mathbf{C} \cdot \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{M}^{-1} \\
&= \mathbf{F}.
\end{aligned} \tag{6.15}$$

As a practical matter, equation (6.14) is more convenient computationally, since the matrix to be inverted has dimension $\mathcal{M} \times \mathcal{M}$, while the matrix in equation (6.9) weighs in at a hefty $\mathcal{N} \times \mathcal{N}$.[1]

Equation (6.12) gives us the entire *a posteriori* probability distribution of $\vec{a}$, *i.e.*, the conditional probability density of $\vec{a}$ given the data and our assumed power spectrum. This probability density is a multivariate Gaussian with mean $\vec{a}_{\rm ML}$ and covariance matrix

$$\mathbf{K} = \left(\mathbf{C}^{-1} + \mathbf{W} \cdot \mathbf{Y}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{Y} \cdot \mathbf{W}\right)^{-1}. \tag{6.16}$$

This probability distribution tells us how likely a particular value of $\vec{a}$ is, given our assumptions. The square roots of the diagonal elements of this matrix can be interpreted as standard error estimates for the elements of $\vec{a}_{\rm ML}$, and those of $\mathbf{Y} \cdot \mathbf{K} \cdot \mathbf{Y}^T$ as error estimates for the elements of $\vec{s}_{\rm ML} = \mathbf{Y} \cdot \vec{a}_{\rm ML}$, which is the maximum-likelihood value of the true CMB signal. In addition, armed with the full covariance matrix $\mathbf{K}$, we can construct constrained realizations of the microwave sky by simply choosing Gaussian random variables from the appropriate distribution.[2] In this way we can construct simulated sky maps with a prob-

---

[1] For the DMR data $\mathcal{N} = 4038$, while $\mathcal{M}$ is only 961 for $l_{\rm max} = 30$. Since matrix inversion is an N^3 process, the reduction in computational time is substantial.

[2] For those interested in trying this at home, I recommend Cholesky-decomposing the covariance matrix $\mathbf{K}$ into a lower-triangular matrix $\mathbf{L}$ times its transpose. Then choose $\vec{q}$ to be an $\mathcal{M}$-dimensional vector of independent Gaussian random variables of zero mean and unit variance, and let $\vec{a} = \vec{a}_{\rm ML} + \mathbf{L} \cdot \vec{q}$.

ability distribution that correctly incorporates both our presumed model and the existing data. These simulated maps can be used to make predictions for other experiments, as we shall see below.

6.3 Accounting for the Monopole and Dipole

Unfortunately, equation (6.1) is not a completely adequate model of the DMR data, because the data do not contain information about all of the multipole moments. In particular, the data are completely insensitive to the monopole contribution, and the dipole is contaminated by the much larger kinematic dipole due to our motion with respect to the CMB rest frame. In this section we will show how the Wiener filter derived above can be modified to account for this. For simplicity of notation, we will ignore beam-smoothing by setting $\mathbf{W} = \mathbf{1}$ throughout this section.

Suppose that the data contain no reliable information about modes with $l \leq l_*$. To account for the monopole and dipole, we should set $l_* = 1$; if we decided to ignore the polemic in Section 2.7 and discard the quadrupole, we would set $l_* = 2$. Then we should modify equation (6.1) as follows:

$$\vec{d} = \vec{s} + \vec{u} + \vec{n} = \mathbf{Y} \cdot \vec{a} + \mathbf{Z} \cdot \vec{b} + \vec{n}, \tag{6.17}$$

where we have excised the elements of $\vec{a}$ and the columns of $\mathbf{Y}$ corresponding to the modes about which we have no information. The dimension of $\vec{a}$ is now $\mathcal{M} = (l_{\max} + 1)^2 - (l_* + 1)^2$ instead $(l_{\max} + 1)^2$, and the corresponding dimensions of $\mathbf{Y}$ and $\mathbf{C}$ are reduced in the same way. $\mathbf{Z}$ is an $\mathcal{N} \times (l_* + 1)^2$ matrix containing the columns removed from $\mathbf{Y}$, $\vec{b}$ is an $(l_* + 1)^2$-dimensional vector containing the unknown coefficients of the excluded modes, and $\vec{u} = \mathbf{Z} \cdot \vec{b}$ is the contribution of these excluded modes to the data.

We can still try to choose our filter $\mathbf{F}$ to be a least-squares estimator in the sense of equation (6.7). However, we must impose an additional constraint on $\mathbf{F}$: since we have no knowledge at all of the true value of $\vec{b}$, we should demand that the result $\mathbf{F} \cdot \vec{d}$ of filtering should be independent of $\vec{b}$. We can do this by imposing the constraint

$$\mathbf{F} \cdot \mathbf{Z} = 0. \tag{6.18}$$

If this constraint is satisfied, then equation (6.8) is still true, and we can solve the constrained minimization problem by the usual method of Lagrange multipliers. The result is

$$(\mathbf{F}-\mathbf{1})\cdot\mathbf{Y}\cdot\mathbf{C}\cdot\mathbf{Y}^T+\mathbf{F}\cdot\mathbf{N}+\boldsymbol{\Lambda}\cdot\mathbf{Z}^T=0, \tag{6.19}$$

where $\boldsymbol{\Lambda}$ is an $\mathcal{N}\times(l_*+1)^2$ matrix of Lagrange multipliers. We can simultaneously solve equations (6.18) and (6.19) to find the filter $\mathbf{F}$:

$$\mathbf{F}=\mathbf{Y}\cdot\mathbf{C}\cdot\mathbf{Y}^T\cdot\mathbf{M}^{-1}(\mathbf{1}-\mathbf{Z}\cdot(\mathbf{Z}^T\cdot\mathbf{M}^{-1}\cdot\mathbf{Z})^{-1}\cdot\mathbf{Z}^T\cdot\mathbf{M}^{-1})\equiv\mathbf{Y}\cdot\mathbf{C}\cdot\mathbf{Y}^T\cdot\mathbf{M}^{-1}\cdot\mathbf{P}. \tag{6.20}$$

This equation differs from the original filter (6.9) only in the projection operator $\mathbf{P}$. This operator simply projects the data vector onto a subspace orthogonal to the unwanted modes, where orthogonality is determined with respect to the inner product $\langle\vec{x},\vec{y}\rangle=\vec{x}^T\cdot\mathbf{M}^{-1}\cdot\vec{y}$.

We shall now show that Wiener filter (6.20) is also a maximum-likelihood filter. Equation (6.11) becomes

$$P(\vec{a},\vec{b}\,|\,\vec{d})\propto\exp\left(-\tfrac{1}{2}\left(\vec{a}^T\cdot\mathbf{C}^{-1}\cdot\vec{a}+(\vec{d}-\mathbf{Y}\cdot\vec{a}-\mathbf{Z}\cdot\vec{b})^T\cdot\mathbf{N}^{-1}\cdot(\vec{d}-\mathbf{Y}\cdot\vec{a}-\mathbf{Z}\cdot\vec{b})\right)\right). \tag{6.21}$$

Completing the square twice, we find that $P(\vec{a},\vec{b}\,|\,\vec{d})\propto\exp\left(-\frac{1}{2}\chi^2\right)$, with

$$\begin{aligned}\chi^2&=(\vec{a}-\vec{a}_{\rm ML})^T\cdot(\mathbf{C}^{-1}+\mathbf{Y}^T\cdot\mathbf{N}^{-1}\cdot\mathbf{Y})\cdot(\vec{a}-\vec{a}_{\rm ML})\\&\quad+(\vec{b}-\vec{b}_{\rm ML})^T\cdot\mathbf{Z}^T\cdot\mathbf{M}^{-1}\cdot\mathbf{Z}\cdot(\vec{b}-\vec{b}_{\rm ML})+\vec{d}^T\cdot\mathbf{M}^{-1}\cdot\vec{d},\end{aligned} \tag{6.22}$$

where

$$\vec{a}_{\rm ML}=\mathbf{C}\cdot\mathbf{Y}^T\cdot\mathbf{M}^{-1}\cdot\mathbf{P}\cdot\vec{d}, \tag{6.23}$$

$$\vec{b}_{\rm ML}=(\mathbf{1}-\mathbf{P})\cdot\vec{d}. \tag{6.24}$$

The maximum-likelihood filter $\mathbf{F}'$ is therefore

$$\mathbf{F}'=\mathbf{Y}\cdot\left(\mathbf{C}^{-1}+\mathbf{Y}^T\cdot\mathbf{N}^{-1}\cdot\mathbf{Y}\right)^{-1}\cdot\mathbf{Y}^T\cdot\mathbf{N}^{-1}\cdot\mathbf{P}, \tag{6.25}$$

which differs from equation (6.14) only by the presence of the projection operator $\mathbf{P}$. We still have $\mathbf{F}=\mathbf{F}'$ exactly as in the last section. Furthermore, equation (6.16) for the covariance matrix $\mathbf{K}$ in the *a posteriori* probability density of $\vec{a}$ is unchanged.

We conclude this section with a pragmatic note. It is in fact not generally necessary to compute the projection operator $\mathbf{P}$ in order to properly project out the unwanted modes. Rather, one can achieve identical results by simply including the unwanted modes together with the others, so that $\vec{a}$ is once again an $(l_{\max}+1)^2$-dimensional vector as it was in the previous section. We extend the signal covariance matrix $\mathbf{C}$ by setting $C_{\mu\nu} = \infty\delta_{\mu\nu}$ in the rows and columns corresponding to the unwanted modes. Then the filter $\mathbf{F}$ and the covariance matrix $\mathbf{K}$ computed in the previous section give precisely the right results.[3]

This result is easy to justify heuristically: since we have no information about the coefficients $\vec{b}$, it is natural to assign them an infinitely wide prior distribution.[4] It is also straightforward to verify by a direct computation that the results of the previous section approach the results of this section as the unknown elements of $\mathbf{C}$ tend to infinity. See [148] for such a calculation.

6.4 Results

In this section we will apply the filter derived above to the two-year DMR data. Figure 6.1 shows the result of applying the Wiener filter to the two-year DMR data. The figure was made from the weighted-average sky map in Figure 2.2 with the usual 20° Galactic cut. The input power spectrum was the standard cold dark matter power spectrum described in Chapter 5. (We will assess the effect of varying the choice of power spectrum below.) It comes as no surprise that there is less information in the Galactic plane.

For illustrative purposes we have also applied the Wiener filter to the world map shown in Figure 2.4. The result is shown in the final panel of the figure. (For this map we used the correct power spectrum as estimated from the raw data shown in the top panel of Figure 2.4. The power spectrum of this data set is approximated well by a broken power law: $C_l \propto l^{-1}$ for $l \leq 5$ and $C_l \propto l^{-2.7}$ for $l > 5$.) Although the large-scale structure in the filtered map is approximately right, it is clear that the Wiener filter does not recover small-scale details correctly. We will see below that the same statement can be made about the real Wiener-filtered data.

The root-mean-square anisotropy level in the filtered map in Figure 6.1 is $30.6\,\mu$K. The ensemble-average expected signal for the power spectrum we have chosen is $42.5\,\mu$K, so we have reason to believe that we have recovered a significant fraction of the power. We

---

[3] One does not of course set anything equal to infinity in one's code: matrices with infinite entries are notoriously ill-conditioned. In practice, if we use equation (6.14) to compute $\mathbf{F}$, we need only $\mathbf{C}^{-1}$, never $\mathbf{C}$, and we can simply set the appropriate entries of $\mathbf{C}^{-1}$ to zero.

[4] In fact, we sneakily did exactly that in equation (6.21), when we did not multiply by $P(\vec{b})$ on the right-hand side.

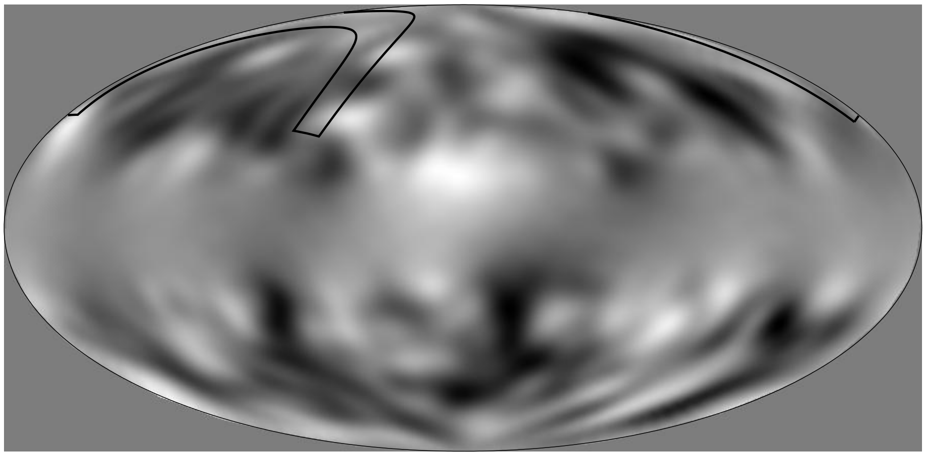

Figure 6.1: The Wiener-filtered DMR data

The result of applying a Wiener filter to the two-year COBE DMR data. The input power spectrum is that of a standard cold dark matter model. See Figures 6.2 and 6.3, as well as the text, for assessments of the uncertainty in the map, and Figure 6.8 for a qualitative assessment of the effect of varying the input power spectrum. The region of sky observed by the Tenerife experiment, *i.e.*, the region with equatorial coordinates $140^\circ < \mathrm{R.A.} < 260^\circ$ and $35^\circ < \delta < 45^\circ$, is also marked on the plot. In Section 6.5 we will use the filtered DMR map to make predictions for this experiment. The part of the strip near 260° in right ascension is split in half in this projection.

can assess the uncertainty in this filtered map in a variety of ways. Given the data and the assumed power spectrum, the *a posteriori* probability distribution of the true signal is Gaussian with mean equal to the filtered signal. The simplest way to assign uncertainties is simply to compute the standard deviation of this probability distribution. In other words, we compute the square roots of the diagonal elements of the covariance matrix $\mathbf{Y}\cdot\mathbf{K}\cdot\mathbf{Y}^T$. These error estimates are plotted in Figure 6.2. The ratio of the signal to the uncertainty is shown in Figure 6.3.

We can get a qualitative idea of the statistical significance of features in the map from these figures: individual pixels in the most prominent hot and cold spots are significant at the three to four sigma level. Of course, the various pixels are not independent: the

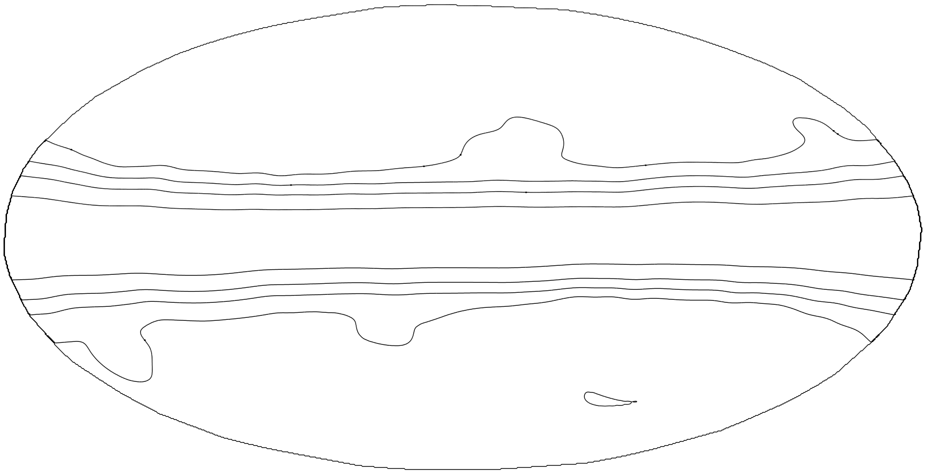

Figure 6.2: Uncertainties in the Wiener-filtered map

We show the one-sigma expected fluctuations about the filtered map in Figure 6.1. The contour levels range from 15 to 35 μK in steps of 0.5 μK. The highest uncertainties are near the Galactic plane, where no data exist.

covariance matrix $\mathbf{Y} \cdot \mathbf{K} \cdot \mathbf{Y}^T$ has off-diagonal elements. Figures 6.4 and 6.5 show the extent of the correlation for one particular pixel. We can also get a handle on the degree of uncertainty in the filtered maps by making constrained realizations of the temperature anisotropy in the manner described in the previous section. Four such realizations are shown in Figure 6.6.

We can define the angular power spectrum of the filtered map:

$$A_l = \frac{1}{2l+1} \sum_{m=-l}^{l} (a_{\mathrm{ML}})^2_{lm}. \tag{6.26}$$

This power spectrum, shown in Figure 6.7, is a useful indicator of the amount of information recovered by the filtering process. On large scales the recovered power A_l is comparable to the input power spectrum, while for modes with high l the power is nearly zero. The Wiener filter, by construction, returns zero values for modes for which we have no information: after all, in the absence of any information from the data, the *a posteriori* probability distribution for a_{lm} is the same as the prior distribution, which is a Gaussian of zero mean.

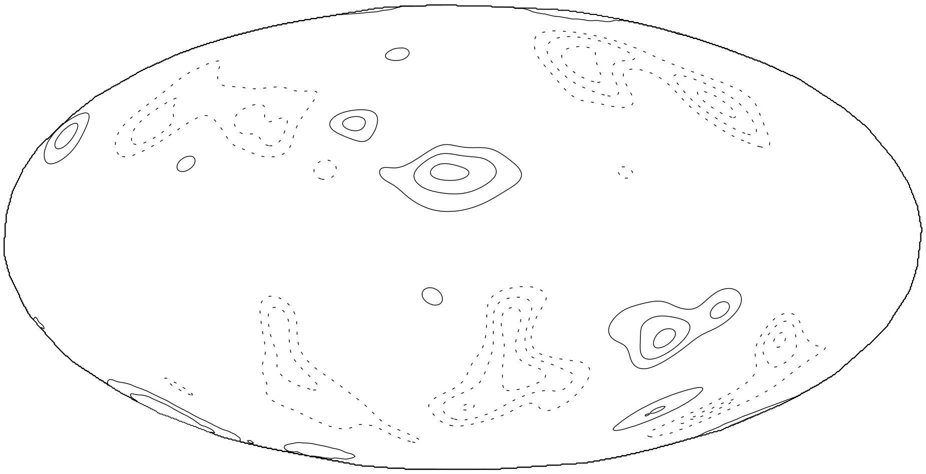

Figure 6.3: Signal-to-noise ratio in the filtered map

The ratio of the filtered map in Figure 6.1 to the uncertainty in Figure 6.2 is plotted. The contours take values $\pm 2, \pm 3, \pm 4$, with dashed lines representing negative values.

Figure 6.8 shows the effect of varying the input power spectrum. The top panel was made with an $n = 0$ power spectrum having very little small-scale power, and the bottom panel is the result of assuming an $n = 2$ spectrum. The large-scale features are seen to be robust, persisting through all of the maps, while the amount of small-scale power varies significantly with the power spectrum.

6.5 Predictions Based on the Wiener-Filtered Map

One use for the Wiener-filtered DMR maps is to make predictions for other experiments. We can process the filtered map through the window function of the other experiment to determine the most probable signal for the experiment to observe, given the DMR data and the assumed power spectrum. Furthermore, since we know the entire conditional probability distribution for ΔT given the DMR data and the assumed prior distribution, we can compute the probability distribution for the predicted signal. Since we have incorpo-

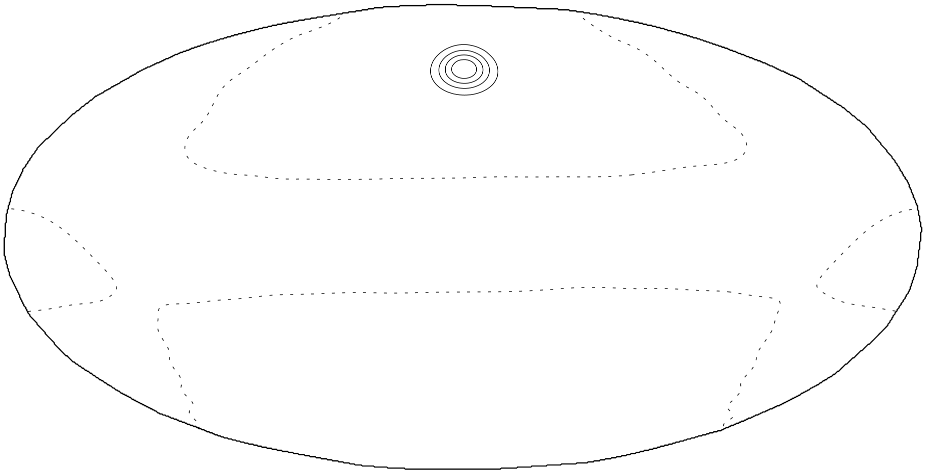

Figure 6.4: Correlated errors in the Wiener-filtered map

We plot the correlation of errors of all points in the Wiener-filtered map shown in Figure 6.1 with respect to the the specific point $(l,b) = (0°, 60°)$. Specifically, if y_i value of the ith pixel of the filtered map and i_0 is the number of the pixel with the above coordinates, then the quantity plotted is $\mathrm{Cov}(y_i, y_{i_0})/\sqrt{\mathrm{Cov}(y_i, y_i)\mathrm{Cov}(y_{i_0}, y_{i_0})}$. The contour levels are 0 (dashed), 0.2, 0.4, 0.6, 0.8. Note that a large portion of the sky, especially that part near the Galactic plane, is weakly anticorrelated with the point i_0. This is a consequence of the fact that the map is constrained to have no monopole contribution.

rated the DMR information, this distribution will in general be narrower than the *a priori* probability distribution associated with our theoretical model. We should therefore be able to place tighter constraints on our model than we could without using the DMR data.

Of course we only expect to gain significantly if we choose an experiment whose window function has significant overlap with the DMR window function, so that the modes that we have estimated well from the DMR data make a substantial contribution to the signal in the other experiment. The Tenerife experiment would seem to be ideal for this purpose, since its beam size of $5^\circ\!.5$ (FWHM) is only slightly smaller than the 7° DMR beam. Unfortunately, the fact that Tenerife is a triple-beam experiment works against us, since the sensitivity to modes with low $|m|$ is quadratically suppressed (see Section 2.4). We will

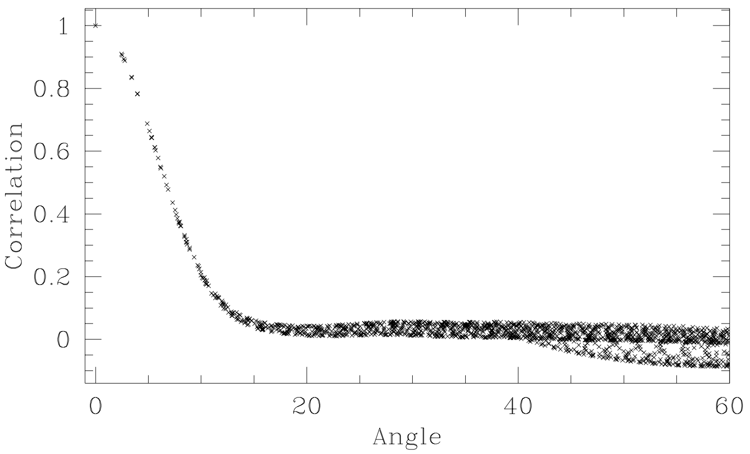

Figure 6.5: Error correlations as a function of angle

This figure shows the correlations shown in Figure 6.4 plotted as a function of angular separation between points i and i_0. The central lobe of the correlation function is fitted well by a Gaussian with standard deviation 5.°7.

see below that this fact greatly reduces the amount of information we can recover from the filtered DMR map.

The Tenerife experiment has taken data covering the range 140° to 260° in right ascension and 35° to 45° in declination. Qualitative maps of the entire data set have been published [146], but only the strip near 40° has been quantitatively analyzed [182]. We will attempt to predict as much as we can of the Tenerife signal over the entire range. The strip of sky observed by the experiment is shown in Figure 6.1.

The process of making predictions is fairly simple. We take the spherical harmonic coefficient $\vec{a}_{\rm ML}$ and multiply it by the Tenerife window function $W^{(\rm Ten)}$ given in Section 2.4.[5] The resulting vector contains the coefficients of a spherical harmonic expansion of the most probable *a posteriori* Tenerife signal $R_{\rm ML}$:

$$R_{\rm ML}(\theta,\varphi)=\sum_{l,m}W_{lm}^{(\rm Ten)}(a_{\rm ML})_{lm}Y_{lm}(\theta,\varphi). \tag{6.27}$$

---

[5] The window function is only simple in a coordinate system in which the chop is in the azimuthal direction. We must either compute $\vec{a}_{\rm ML}$ in these coordinates in the first place or else rotate to them afterwards. The Tenerife experiment always chops in right ascension, so the rotation is easy to perform.

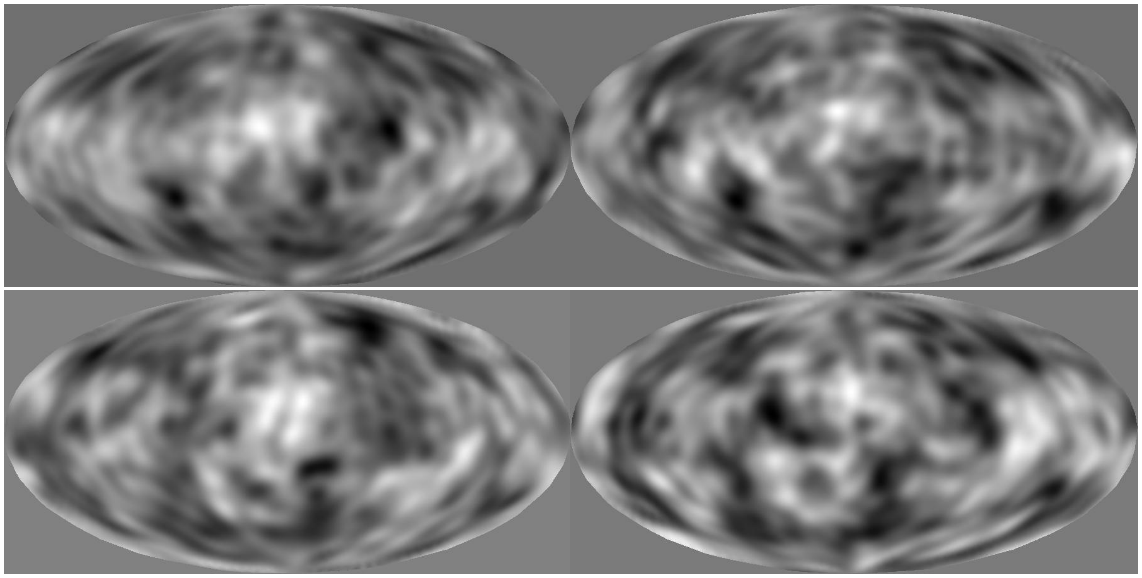

Figure 6.6: Constrained realizations

Four constrained realizations of the microwave sky are shown. The realizations represent typical expected fluctuations about the Wiener-filtered map in Figure 6.1, and were drawn from the Gaussian probability distribution described in the previous section.

The result is shown in Figure 6.9. The portion of sky plotted corresponds to the region of sky covered by the Tenerife experiment. The root-mean-square value of $R_{\rm ML}$ over this region is $16.3\,\mu$K. For the standard CDM model we are considering, the expected r.m.s. signal is $32.7\,\mu$K, and so it appears that we have predicted a significant fraction of the total signal. Recall that the Wiener filter returns low values for modes for which it has little information, and values near the expected value for modes with high signal-to-noise ratio (see Figure 6.7). The ratio of the Wiener-filtered prediction to the *a priori* expectation value is thus a rough indicator of how much signal has been recovered.

We can of course do much better than this qualitative argument. We know that the vector $\vec{a}$ is Gaussian with mean $\vec{a}_{\rm ML}$ and covariances given by (6.16). The Tenerife response R is therefore Gaussian distributed with mean $R_{\rm ML}$ and covariance

$$\mathrm{Cov}(R(\hat{\mathbf{r}}_1),R(\hat{\mathbf{r}}_2))=\sum_{\mu,\nu}W_\mu^{(\mathrm{Ten})}Y_\mu(\hat{\mathbf{r}}_1)W_\nu^{(\mathrm{Ten})}Y_\nu(\hat{\mathbf{r}}_2)K_{\mu\nu}, \tag{6.28}$$

where μ and ν represent pairs of indices (lm) as usual. The standard error associated with

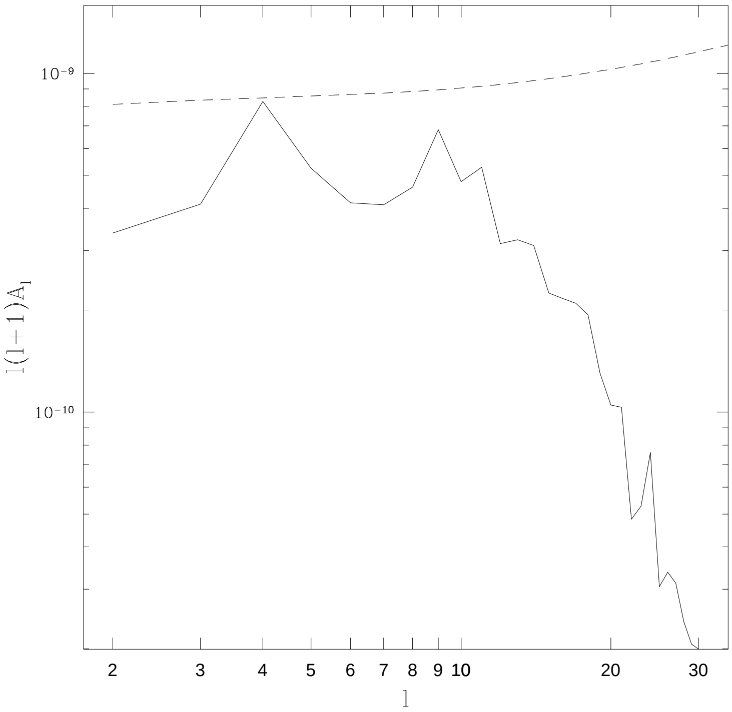

Figure 6.7: Power spectrum of the Wiener-filtered map

The power spectrum A_l defined in equation (6.26) is shown. The standard CDM input power spectrum is also plotted for comparison.

each point in Figure 6.9 is therefore $\sigma(\hat{\mathbf{r}}) = \sqrt{\mathrm{Cov}(R(\hat{\mathbf{r}}), R(\hat{\mathbf{r}}))}$. The standard error ranges from 26.4 μK to 29.7 μK.

One way to get an idea of the amount of information we have gained by this exercise is to compare these standard errors with the level of uncertainty we would have without the DMR data. If we made no use of the DMR data, then our probability distribution for $R(\hat{\mathbf{r}})$ would simply be Gaussian with zero mean and standard error given by the r.m.s. expected signal of 32.7 μK. Using the DMR data reduces the standard error by about 15%. This may not seem like anything to write home about, but it should amount to at least a modest improvement in our ability to assess the consistency of this model with the Tenerife data. Furthermore, the large positive feature near right ascension 250° appears to have moderately high statistical significance.

In Figure 6.11 we have plotted the predicted signal and associated errors in several

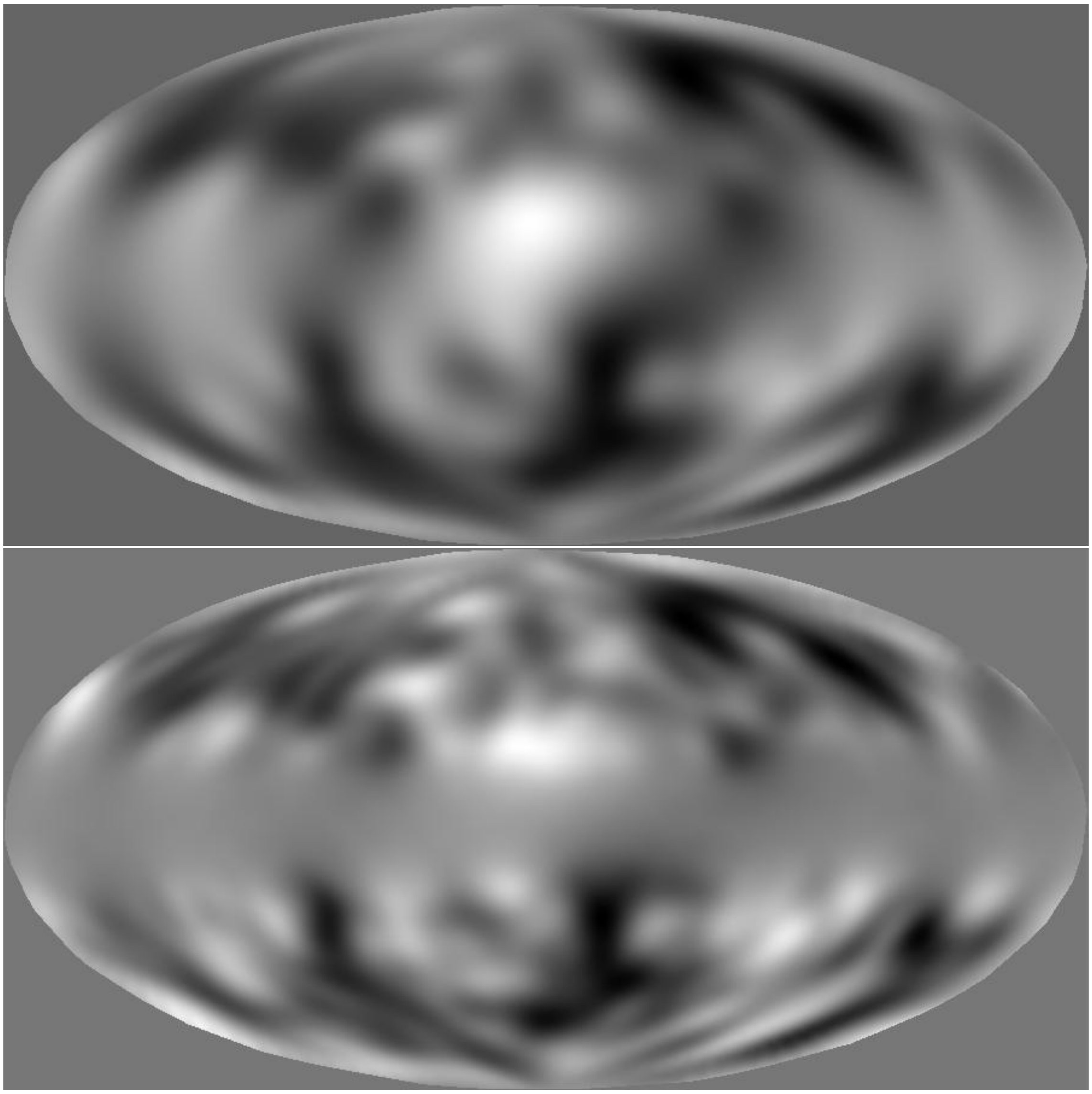

Figure 6.8: Wiener filtered maps with different power spectra

The maps shown here were produced by Wiener filtering in the same manner as the map in Figure 6.1, but with different power spectra. The top panel was made with an $n = 0$ Sachs-Wolfe power spectrum, using the formula in equation (2.29). The bottom panel was made with an $n = 2$ spectrum. Note that the large features in all three of the Wiener filtered maps coincide, but the amount of small-scale power varies with the input power spectrum.

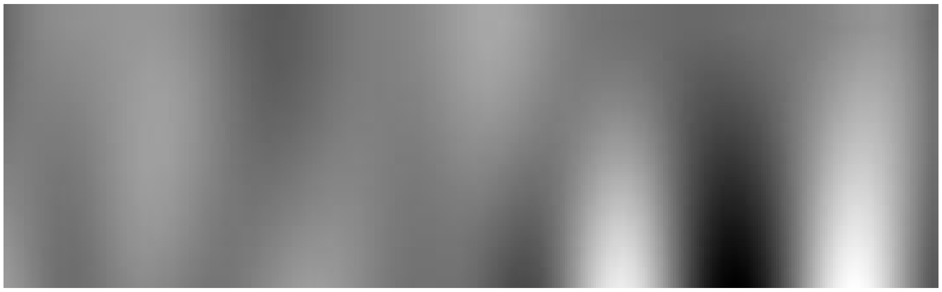

Figure 6.9: Tenerife predictions

A map of the most probable (*a posteriori*) Tenerife signal over the region $120^\circ <$ RA $< 260^\circ$, $35^\circ < \delta < 45^\circ$. The greyscale ranges from $-60\,\mu$K to $60\,\mu$K, and the uncertainty associated with each point is approximately $28\,\mu$K.

individual strips for greater ease of comparison with the actual data. The maximum at right ascension of 250° coincides almost exactly with the largest peak in the Tenerife map [146].

The standard errors simply represent the diagonal part of the covariances in equation (6.28). For a complete statistical description of the prediction, we need to know the correlations between distinct points as well. Because of the beam-switching strategy, these correlations can be quite important. These correlations are shown in Figure 6.10 for the central point in the map. The general shape of the correlations is as one would expect from the beam-switching patterns: a positive lobe flanked by two symmetric negative lobes. The following fitting formula reproduces the three main lobes to about 5% accuracy, although it misses the outer ripples:

$$\begin{aligned}\mathrm{Cov}(R(\hat{\mathbf{r}}_1)R(\hat{\mathbf{r}}_2)) &= \mathrm{Cov}(R(\hat{\mathbf{r}}_1)R(\hat{\mathbf{r}}_1))e^{-\Delta\delta\theta^2/2(4^\circ\!.7)^2}\times\\ &\left(e^{-\Delta A^2/2(3^\circ\!.0)^2}-0.69e^{-(\Delta A-9^\circ\!.2)^2/2(3^\circ\!.0)^2}-0.69e^{-(\Delta A+9^\circ\!.2)^2/2(3^\circ\!.0)^2}\right),\end{aligned}\tag{6.29}$$

where $\Delta\delta$ is the declination difference between $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$, and $\Delta A = \Delta\varphi\cos\delta_1$ is the separation in azimuth.

As with the original Wiener-filtered map, it may be helpful to look at constrained realizations of the statistical fluctuations about the maximum-likelihood Tenerife prediction in order to get some idea of the level of uncertainty. Five such realizations are plotted in Figure 6.12. We show both greyscale maps for comparison with Figure 6.9 and linear plots of the strip at declination 35° for comparison with Figure 6.11.

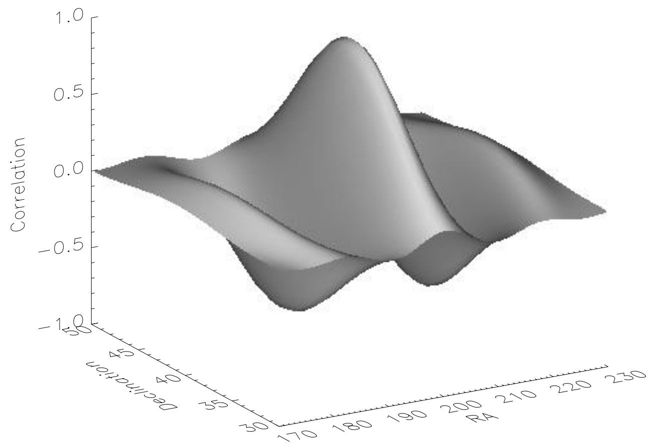

Figure 6.10: Correlations in the Tenerife predictions

This figure shows the correlations in the residuals of the various points in the map with respect to the central point. The correlation is divided by the variance of the central point so that its maximum value is unity.

6.6 Conclusions

Wiener filtering is a promising tool for the analysis of CMB sky maps. Filtering provides a significant improvement in the signal-to-noise ratio in the regions covered by the raw data, and allows some small amount of information to be reconstructed about the anisotropy within the Galactic cut region. It is possible to identify several hot and cold spots in the map which carry high statistical significance.

This reduction in noise is not without a price. In order to apply the Wiener filter, one needs to assume a power spectrum. The cleaned data therefore depend on more assumptions than do the raw data. However, the maps do not undergo great qualitative changes as one varies the slope of the power spectrum over a wide range of reasonable values.

One great advantage of the Wiener-filtered map is that the statistical properties of the residuals are known precisely, assuming the underlying model is correct. It is therefore possible to assess the goodness of fit of any other data set with the combined DMR data and assumed theory. In addition, it is easy to make constrained realizations of the microwave sky with the correct *a posteriori* probability distribution. This technique can in principle

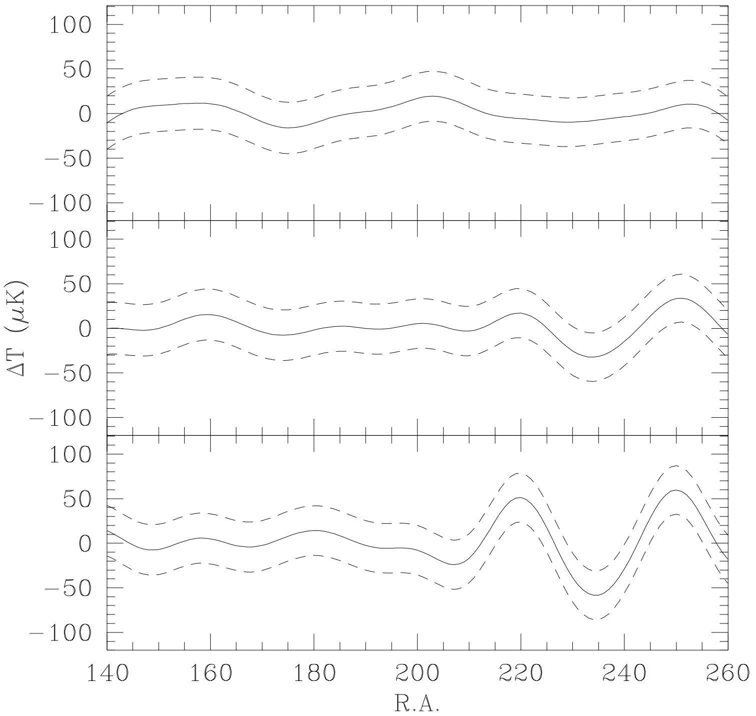

Figure 6.11: Tenerife predictions with uncertainties

The solid lines show slices through the map in Figure 6.9, and the dashed lines show plus and minus one-sigma errors. The top, middle,and bottom panels correspond to declinations of 45°, 40°, and 35° respectively.

be used to perform Monte-Carlo simulations of the CMB for comparison with experiments on similar angular scales.

The Tenerife experiment seems to be the most promising experiment on which to test the possibility of using the Wiener filtered maps to make predictions. Unfortunately, the triple-beam strategy used in the experiment drastically reduces Tenerife's sensitivity to the modes that are constrained well by the DMR. It is possible, however, to predict some modest fraction of the Tenerife signal from the filtered DMR data. By incorporating this information into a Tenerife analysis, it should be possible to place somewhat tighter constraints on models than would be possible with the Tenerife data alone.

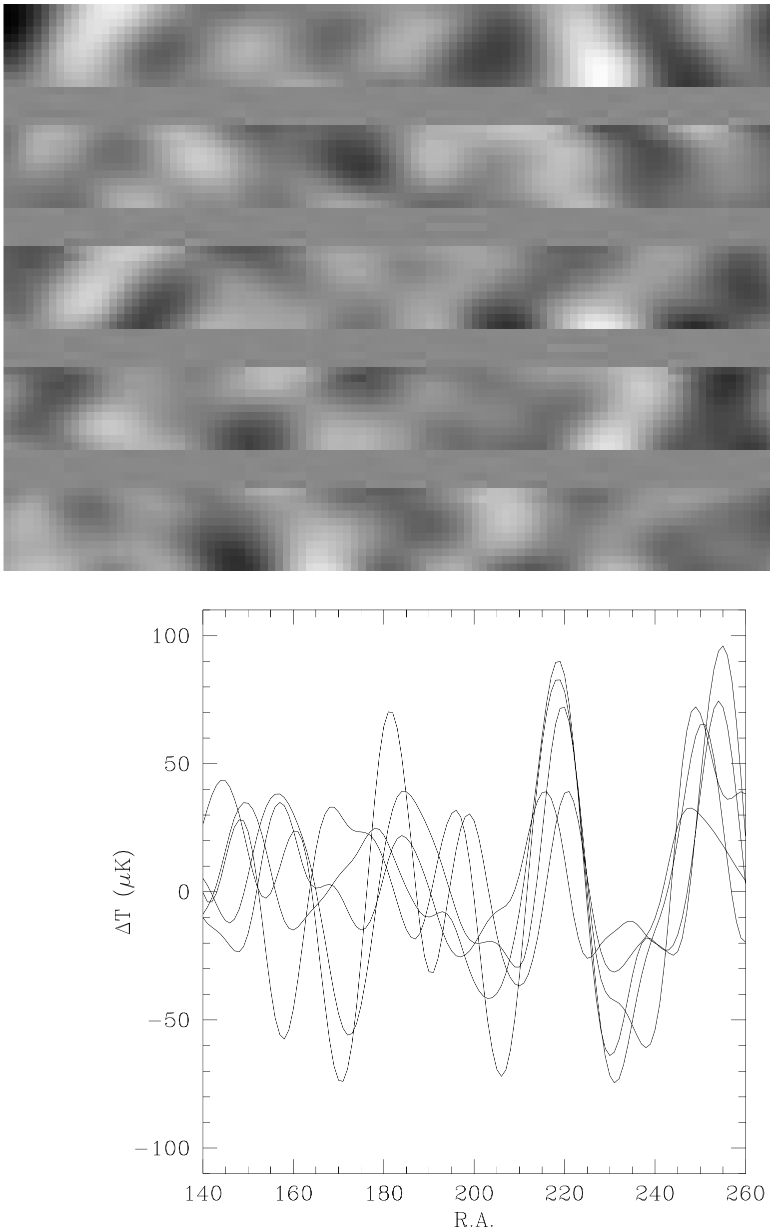

Figure 6.12: Constrained realizations of the Tenerife data

We show five constrained realizations of the Tenerife predictions to assist the reader in qualitatively assessing the uncertainties in the predictions in Figures 6.9 and 6.11. The top panel shows five realizations covering the same range as Figure 6.9. The greyscale in this map ranges from -110 to 100 μK. The bottom panel is a strip through the five realizations at declination 35°, and is thus to be compared with the lowest panel of Figure 6.11.

Bibliography

[1] Abbott, L.F. & Schaefer, R.K. 1986, Ap. J., 308, 546.

[2] Abbott, L.F. & Wise, M.B. 1984, Ap. J. Lett., 282, 47.

[3] Abramowitz, M. & Stegun, I.A., eds., 1965, *Handbook of Mathematical Functions* (Dover, New York).

[4] Adams, F.C., Bond, J.R., Freese, K., Freeman, J.A. & Olinto, A. 1993, Phys. Rev. D, 47, 426.

[5] Adler, R.J. 1981, *The Geometry of Random Fields* (J. Wiley, New York).

[6] Banday, A.J. *et al.* 1994, Ap. J. Lett., 436, 99.

[7] Barrow, J.D., Juszkiewicz, R., & Sonoda, D.H. 1985, M.N.R.A.S., 213, 917.

[8] Bardeen, J.M., Steinhardt, P.J., & Turner, M.S. 1983, Phys. Rev. D., 28, 679.

[9] Bartlett, J.G., Blanchard, A., Silk, J., & Turner, M.S. 1995, Science, 267, 980.

[10] Beichman, C.A. *et al.* 1988, *IRAS Explanatory Supplement*, Volume 1, page X-32 (NASA, Washington, DC).

[11] Bennett, C.L. *et al.* 1992, Ap. J., 391, 466.

[12] Bennett, C.L. *et al.* 1992, Ap. J. Lett., 396, 7.

[13] Bennett, C.L. *et al.* 1994, Ap. J., 436, 423.

[14] Bennett, D.P., Stebbins, A. & Bouchet, F.R. 1992, Ap. J. Lett. 399, 5.

[15] Berger, J.O. 1985, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York).

[16] Bertschinger, E. *et al.* 1990, Ap. J., 364, 370.

[17] Birkinshaw, M. & Hughes, J.P. 1994, Ap. J., 420, 33.

[18] Boggess, N.W. *et al.* 1992, Ap. J., 397, 420.

[19] Bond, J.R. 1994, in *Proceedings of the IUCAA Dedication Ceremonies*, ed. T. Padmanabhan (John Wiley & Sons, New York).

[20] Bond, J.R. 1994a, Astrophys. Lett. and Comm., in press.

[21] Bond, J.R. 1994b, preprint (CITA–94-27, astro-ph–9407044).

[22] Bond J.R., Crittenden R., Davis R.L., Efstathiou G. & Steinhardt P.J. 1994, Phys. Rev. Lett., 72, 13.

[23] Bond, J.R. 1995, "Theory and Observations of the Cosmic Background Radiation," to appear in *Cosmology and Large-Scale Structure*, proceedings of the Les Houches Summer School, Session LX, R. Schaeffer, ed. (Elsevier Science Publishers, Netherlands).

[24] Bond, J.R. & Efstathiou, G. 1984, Ap. J. Lett., 285, 45.

[25] Bond, J.R. & Efstathiou, G. 1987, M.N.R.A.S., 226, 655.

[26] Brandenberger, R.H. 1993, in *Current Topics in Astrofundamental Physics, Second International School of Physics 'D Chalonge,'*, N. Sanchez & A. Zichichi, eds., p. 272 (World Scientific, Singapore).

[27] Brandenberger, R.H. & Turok, N. 1986, Phys. Rev. D, 33, 2182.

[28] Brandt, W.N. *et al.* 1994, Ap. J., 424, 1.

[29] Bucher, M., Goldhaber, A., & Turok, N. 1995, Phys. Rev., in press.

[30] Bunn, E.F., Hoffman, Y., & Silk, J. 1994, Ap. J., 425, 359.

[31] Bunn, E.F., Scott, D., & White, M. 1995, Ap. J. Lett, 441, 9.

[32] Bunn, E.F. & Sugiyama, N. 1995, Ap. J., in press.

[33] Bunn, E.F., White, M., Srednicki, M., & Scott, D. 1994, Ap. J., 429, 1.

[34] Bunn, E.F., Fisher, K.B., Hoffman, Y., Lahav, O., Silk, J., & Zaroubi, S. 1994, Ap. J. Lett., 432, 75.

[35] Cæsar, G.J. 55 B.C., *De Bello Gallico.*

[36] Carroll, S.M., Press, W.H., & Turner, E.L. 1992, Ann. Rev. Astron. Astrophys., 30, 499.

[37] Chambers, K.C. & McCarthy, P.J. 1990, Ap. J. Lett., 354, 9.

[38] Clapp, A.C. *et al.* 1994, Ap. J. Lett., 433, 57.

[39] Copernicus, N. 1543, *De Revolutionibus.*

[40] Corey, B.E. & Wilkinson, D.T. 1976, Bull. Am. Astron. Soc., 8, 351.

[41] Coulson, D., Ferreira, P., Graham, P. & Turok, N. 1994, Nature, 368, 27.

[42] Cowie, L.L. 1991, in *Observational Tests of Cosmological Inflation*, T. Shanks *et al.*, eds., p. 257 (Kluwer, Dordrecht).

[43] Crane, S. 1899, *War is Kind and Other Lines* (F.A. Stokes Co., New York).

[44] Crittenden, R., Bond, J.R., Davis, R.L., Efstathiou, G. & Steinhardt, P.J. 1993, Phys. Rev. Lett., 71, 324.

[45] Davies, R.D. *et al.* 1992, M.N.R.A.S., 258, 605.

[46] Davis, M. & Peebles, P.J.E. 1983, Ap. J., 267, 465.

[47] Davis, M., Summers, F.J., & Schlegel, D., 1992, Nature, 359, 393.

[48] Davis, R.L., Hodges, H.M., Smoot, G.F., Steinhardt, P.J., & Turner, M.S., 1992, Phys. Rev. Lett., 69, 1856.

[49] Dekel, A. 1994, Ann. Rev. Astron. Astrophys., 32, 371.

[50] de Oliveira Costa, A. & Smoot, G.F., COBE preprint (astro-ph/9412003).

[51] Devlin, M., *et al.* 1994, Ap. J. Lett., 430, 1.

[52] Dicke, R.H., Peebles, P.J.E., Roll, P.G. & Wilkinson, D.T. 1965, Ap.J., 142, 414.

[53] Dodelson, S. & Jubas, J. 1993, Phys. Rev. Lett., 70, 2224.

[54] Doyle, A.C. 1927, "The Adventure of the Blanched Soldier," in *The Case Book of Sherlock Holmes* (George H. Doran, New York).

[55] Dressler, A., Oemler, A., Sparks, W.B. & Lucas, R.A. 1994, Ap. J. Lett., 435, 23.

[56] Durrer, R., Howard, A. & Zhou, Z.-H. 1994, Phys. Rev. D, 49, 681.

[57] Efstathiou, G. 1989, "Cosmological Perturbations," in Proceedings of the 36th Scottish Universities Summer School in Physics, Edinburgh, July 24– August 11, 1989 (IOP pub.).

[58] Efstathiou, G. 1995, "Observations of Large-Scale Structure," to appear in *Cosmology and Large-Scale Structure*, proceedings of the Les Houches Summer School, Session LX, R. Schaeffer, ed. (Elsevier Science Publishers, Netherlands).

[59] Efstathiou, G., Bond, J.R. & White, S.D.M. 1992, M.N.R.A.S., 258, 1p.

[60] Efstathiou, G., Sutherland, W.J., & Maddox, S.J. 1990, Nature, 348, 705.

[61] Einstein, A. 1922, Ann. Phys., 69, 436.

[62] Euripides 410 B.C., *Ion*, translated by Robert Potter.

[63] Friedmann, A. 1922, Zeitschrift für Physik, 10, 377.

[64] Gaier, T., *et al.* 1992, Ap. J. Lett., 398, 1.

[65] Gallot, S., Hulin, D. & Lafontaine, J. 1987 *Riemannian Geometry* (Springer-Verlag, New York).

[66] Ganga, K., Cheng, E., Mayer, S., & Page, L. 1993, Ap. J. Lett., 410, 57.

[67] Ganon, G. & Hoffman, Y. 1993, Ap. J. Lett., 415, 5.

[68] Górski, K. 1994, Ap. J. Lett., 430, 85.

[69] Górski, K. 1995, private communication.

[70] Górski, K., Stompor, R., & Juszkiewicz, R. 1993, Ap. J. Lett., 410, 1.

[71] Górski, K. *et al.* 1994, Ap. J. Lett., 430, 89.

[72] Górski, K., Ratra, B., Sugiyama, N. & Banday, A. 1995, COBE preprint (astro-ph/9502034).

[73] Górski, K. *et al.* 1995, in preparation.

[74] Gott, J.R. 1980, M.N.R.A.S., 193, 153.

[75] Gould, A. 1993, Ap. J. Lett., 403, 51.

[76] Gundersen, J.O. *et al.* 1993, Ap. J. Lett., 413, 1.

[77] Guth, A. & Pi, S.-Y. 1982, Phys. Rev. Lett., 49, 1110.

[78] Halley, E. 1720, Phil. Trans. Roy. Soc. London, 31, 22.

[79] Hancock, S. *et al.* 1994, Nature, 367, 333.

[80] Harrison, E.R. 1970, Phys. Rev. D, 1, 2726.

[81] Hauser, M.G. & Peebles, P.J.E. 1973, Ap. J., 185, 757.

[82] Hinshaw, G., Kogut, A., Gorski, K.M., Banday, A.J., Bennett, C.L., Lineweaver, C., Lubin, P., Smoot, G.F., & Wright, E.L. 1994, Ap. J., 431, 1.

[83] Hoffman, Y. & Ribak, E. 1991, Ap. J. Lett. 380, 5.

[84] Holtzman, J. A. 1989, Ap. J. Supp., 71, 1.

[85] Hu, W. 1995, *Wandering in the Background: A Cosmic Background Explorer*, Ph.D. Dissertation, U.C. Berkeley Physics Department.

[86] Hu, W., Bunn, E.F., & Sugiyama, N. 1995, Ap. J., in press.

[87] Hu, W., Scott, D. & Silk, J. 1994, Phys. Rev. D, 49, 648.

[88] Hu, W. & Sugiyama, N. 1995, Ap. J., 444, 489.

[89] Hu, W. & Sugiyama, N. 1995, Phys. Rev. D, 51, 2599.

[90] Hu, W., Sugiyama, N. & Silk, J. 1995, preprint (CfPA-TH-95-08, astro-ph/9504057).

[91] Hu, W., Scott, D., Sugiyama, N. & White, M. 1995, preprint (CfPA-TH-95-10, UTAP-208, astro-ph/9505043).

[92] Hubble, E.P. 1929, Proc. Natl. Acad. Sci. U.S. 15, 169.

[93] Jackson, J.D. 1975, *Classical Electrodynamics* (John Wiley & Sons, New York).

[94] Jackson, P.D. *et al.* 1991, in P.A.S.P. Conference Series 25, *Astronomical Data Analysis Software and Systems I*, D.M. Worrall, C. Biemesderfer & J. Barnes, eds. (A.S.P., San Francisco).

[95] Jacoby, G.H. *et al.* 1992, Pub. Astron. Soc. Pacific, 104, 599.

[96] Jing, Y.-P. & Fang, L.-Z. 1994, Phys. Rev. Lett., 73, 1882.

[97] Jones, M. *et al.* 1993, Nature, 365, 320.

[98] Kaiser, N. & Stebbins, A. 1984, Nature, 310, 391.

[99] Karhunen, K. 1947, *Über lineare Methoden in der Wahrschleinlichkeitsrechnung* (Kirjapaino oy. sana, Helsinki).

[100] Klypin, A.A., Strukov, I.A. & Skulachev, D.P. 1992, M.N.R.A.S., 258, 71.

[101] Kochanek, C.S. 1993, Ap. J., 419, 12.

[102] Kochanek, C.S. 1994, in *Critique of the Sources of Dark Matter in the Universe*, in press.

[103] Kofman, L., Gnedin, N. & Bahcall, N. 1993, Ap. J., 413, 1.

[104] Kofman, L.A. & Starobinski, A.A. 1985, Sov. Astron. Lett., 11, 271.

[105] Kogut, A. *et al.* 1992, Ap. J., 401, 1.

[106] Kogut, A. *et al.* 1993, Ap. J., 419, 1.

[107] Kogut, A. *et al.* 1995, Ap. J. Lett., 439, 29.

[108] Kolb, E.W. & Turner, M.S. 1990, *The Early Universe*, Frontiers in Physics vol. 69 (Addison-Wesley, New York).

[109] Kolb, E.W. & Vadas, S. 1993, Phys. Rev. D, 50, 2479.

[110] Lahav, O., Fisher, K.B., Hoffman, Y., Scharf, C.A., & Zaroubi, S. 1994, Ap. J. Lett., 423, 93.

[111] Liddle, A.R. & Lyth, D.H. 1992, Phys. Lett. B, 291, 391.

[112] Liddle, A.R. & Lyth, D.H. 1993, Phys. Rep., 231, 1.

[113] Liddle, A.R. & Lyth, D.H. 1995, M.N.R.A.S., in press.

[114] Lifshitz, E.M. & Khalatnikov, I.M. 1963, Adv. Phys., 12, 185.

[115] Linde, A. 1990, *Particle Physics and Inflationary Cosmology* (Harwood Academic).

[116] Lineweaver, C.H. 1994, *Correlation Function Analysis of the COBE Differential Microwave Radiometer Sky Maps*, Ph.D. Dissertation, Physics Department, U.C. Berkeley.

[117] Lineweaver, C.H., Smoot, G.F., Bennett, C.L., Wright, E.L., Tenorio, L., Kogut, A., Keegstra, P.B., Hinshaw, G., & Banday, A.J. 1994, Ap.J., 436, 452.

[118] Lobell, F. 1931, Ber. Verhand. Sachs. Akad. Wiss. Leipzig Math. Phys. KL, 83, 167.

[119] Longair, M.S. 1993, Q.J.R.A.S., 34, 157.

[120] Loveday, J., Efstathiou, G., Peterson, B.A., & Maddox, S.J. 1992, Ap. J. Lett., 400, 43.

[121] Lyth, D.H. & Stewart, E. 1990, Phys. Lett. B, 252, 336.

[122] Lyth, D.H. & Woszczyna, A. 1995, preprint (astro-ph/9501044).

[123] Mather, J.C. *et al.* 1994, Ap. J., 420, 439.

[124] Meinhold, P. *et al.* 1993, Ap. J. Lett., 409, 1.

[125] Meyer, S.S., Cheng, E.S., & Page, L. 1991, Ap. J. Lett., 371, 7.

[126] Mukhanov, V.F., Feldman, H.A., & Brandenberger, R.H. 1992, Phys. Rep., 215, 203.

[127] Narlikar, J.V. & Padmanabhan, T. 1991, Annu. Rev. Astron. Astrophys., 29, 325.

[128] *NAG Fortran Library*, Numerical Algorithms Group, 256 Banbury Road, Oxford, U.K.

[129] Olive, K. 1990, Phys. Rep., 190, 307.

[130] Olbers, H.W.M. 1826, "Über die Durchsichtigkcit des Weltraumes," *Bode Jahrbuch*, 110.

[131] Ostriker, J. P. 1993, Ann. Rev. Astron. Astrophys., 31, 689.

[132] Padmanabhan, T. 1993, *Structure Formation in the Universe* (Cambridge University Press, Cambridge)

[133] Peacock, J. & Dodds, S. 1994, M.N.R.A.S., 267, 1020.

[134] Peebles, P.J.E. 1971, *Physical Cosmology* (Princeton University Press, Princeton).

[135] Peebles, P.J.E. 1973, Ap. J., 185, 413.

[136] Peebles, P.J.E. 1980, *The Large-Scale Structure of the Universe* (Princeton University Press, Princeton).

[137] Peebles, P.J.E. 1982, Ap. J. Lett., 263, 1.

[138] Peebles, P.J.E. 1993, *Principles of Physical Cosmology* (Princeton University Press, Princeton).

[139] Peebles, P.J.E., Schramm, D.N., Turner, E.L., & Kron, R.G. 1991, Nature, 352, 769.

[140] Peebles, P.J.E. & Yu, J.T. 1970, Ap. J., 162, 815.

[141] Penzias, A.A. & Wilson, R.W. 1965, Ap. J., 142, 419.

[142] Poe, E.A. 1848, *Eureka: a Prose Poem* (G.P. Putnam, New York).

[143] Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. 1992, *Numerical Recipes in Fortran*, 2nd ed. (Cambridge University Press, Cambridge, U.K.).

[144] Ratra, B. & Peebles, P.J.E. 1994, Ap. J. Lett., 432, 5.

[145] Readhead, A.C.S. & Lawrence, C.R. 1992, Annu. Rev. Astron. Astrophys., 30, 653.

[146] Rebolo, R., Watson, R.A., Gutierrez de la Cruz, C.M., Davies, R.D., Lasenby, A.N. & Hancock, S. 1994, Astrophys. Lett. & Commun., in press.

[147] Rombauer, I.S. & Becker, M.R. 1975, *The Joy of Cooking* (Bobbs-Merrill, Indianapolis).

[148] Rybicki, G.B. & Press, W.H. 1992, Ap. J. 398, 169.

[149] Sachs, R.K. & Wolfe, A.M. 1967, Ap. J., 147, 73.

[150] Scaramella, R. 1993, Ap. J., 411, 1.

[151] Scaramella, R. & Vittorio, N. 1993, M.N.R.A.S., 263, L17.

[152] Scharf, C., Hoffman, Y., Lahav, O. & Lynden-Bell, D. 1992, M.N.R.A.S., 256, 229.

[153] Scott, D., Silk, J. & White, M. 1994, preprint.

[154] Scott, D., Srednicki, M. & White, M. 1994, Ap. J. Lett. 421, 5.

[155] Seljak, U. & Bertschinger, E. 1993, Ap. J. Lett., 417, 9.

[156] Shakespeare, W. 1579, *A Comedy of Errors.*

[157] Silk, J. 1967, Nature, 215, 1155.

[158] Silk, J. 1968, Ap. J., 151, 459.

[159] Smith, M.S., Kawano, L.H. & Malaney, R.A. 1993, Ap. J. Supp., 85, 219.

[160] Smoot, G.F., Gorenstein, M.V. & Muller, R.A. 1977, Phys. Rev. Lett., 39, 898.

[161] Smoot, G.F. *et al.* 1990, Ap. J., 360, 685.

[162] Smoot, G.F. *et al.* 1991, Ap. J. Lett., 371, 1.

[163] Smoot, G.F. *et al.* 1992, Ap. J. Lett., 396, 1.

[164] Smoot, G.F., Tenorio, L., Banday, A.J., Kogut, A., Wright, E.L., Hinshaw, G., & Bennett, C.L. 1994, Ap. J., 437, 1.

[165] Spivak, M. 1979, *A Comprehensive Introduction to Differential Geometry* (Publish or Perish, Boston).

[166] Srednicki, M., White, M., Scott, D. & Bunn, E.F. 1993, Phys. Rev. Lett., 71, 3747.

[167] Stark, P.B. 1993, Ap J. Lett., 408, 73.

[168] Stevens, D., Scott, D, & Silk, J. 1993, Phys. Rev. Lett., 71, 20.

[169] Sugiyama, N. 1995, preprint (astro-ph/9412025).

[170] Sugiyama, N. & Gouda, N. 1992, Prog. Theor. Phys., 88, 803.

[171] Sugiyama, N. & Silk, J. 1994, Phys. Rev. Lett., 73, 509.

[172] Sunyaev, R.A, & Zel'dovich, Ya.B. 1970, Astrophys. Space Sci., 7, 3.

[173] Tegmark, M., preprint (astro-ph/9412064).

[174] Tegmark, M. & Bunn, E.F. 1995, Berkeley preprint (astro-ph/9412005).

[175] Tegmark, M., Bunn, E.F., & Hu. W. 1994, Ap.J., 434, 1.

[176] Thierren, C.W. 1992, *Discrete Random Signals and Statistical Signal Processing* (Prentice-Hall).

[177] Torres, S. 1994, Ap. J. Lett., 423, 9.

[178] Turner, M.S., White, M. & Lidsey, J.E. 1993, Phys. Rev. D, 48, 4613.

[179] Turok, N. 1991, Physica Scripta, T36, 135.

[180] Turok, N. & Spergel, D.N. 1991, Phys. Rev. Lett., 66, 3093.

[181] Walker, P.N., Steigman, G., Schramm, D.N., Olive, K.A. & Kang, H.-S. 1991, Ap. J., 376, 51.

[182] Watson, R.A. *et al.* 1992, Nature, 357, 660.

[183] Watson, R.A. & Gutierrez de la Cruz, C.M. 1993, Ap. J. Lett. 419, 5.

[184] White, M. 1994, Astron. & Astrophys., 290, L1.

[185] White, M. & Bunn, E.F. 1995, Ap. J., in press.

[186] White, M. & Bunn, E.F. 1995, Ap. J. Lett., in press.

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

[188] White, M., Scott, D. & Silk, J. 1994, Ann. Rev. Astron. Astrophys., 32, 319.

[189] White, M. & Srednicki, M. 1995, Ap. J., in press (preprint astro-ph/9402037).

[190] White, S.D.M. 1995, "Galaxy Formation," to appear in *Cosmology and Large-Scale Structure*, proceedings of the Les Houches Summer School, Session LX, R. Schaeffer, ed. (Elsevier Science Publishers, Netherlands).

[191] White, S.D.M., Efstathiou, G. & Frenk C. 1993, M.N.R.A.S., 262, 1023.

[192] Wilbanks, T.M. *et al.* 1994, Ap. J. Lett., 427, 75.

[193] Wolf, J.A. 1974, *Spaces of Constant Curvature* (Publish or Perish, Boston).

[194] Wright, E.L. *et al.* 1992, Ap. J. Lett., 396, 13.

[195] Wright, E.L. *et al.* 1994, Ap. J., 420, 1.

[196] Wright, E.L., Smoot, G.F., Bennett, C.L. & Lubin, P.M. 1994, Ap. J., 436, 443.

[197] Yamamoto, K., Sasaki, M. & Tanaka, T. 1995, Kyoto preprint.

[198] Zaroubi, S., Hoffman, Y., Fisher, K.B., Lahav, O., & Lynden-Bell, D. 1994, preprint (astro-ph/9410080).

[199] Zel'dovich, Ya.B. 1965, Adv. Astron. Astrophys., 3, 241.

Appendix A

The Aitoff Projection

The Aitoff projection is a common equal-area projection of the sphere into the plane. We make frequent use of it in this dissertation, and we therefore include the equations of the projection here. These formulae are taken from the IRAS Explanatory Supplement [10].

For the forward projection, in which a point on the sphere with spherical coordinates (θ, φ) is mapped into a point with Cartesian coordinates (x, y) in the plane, we define ρ and γ by

$$\rho = \cos^{-1}\left(\sin\theta\cos(\varphi/2)\right), \tag{A.1}$$

$$\gamma = \sin^{-1}\left(\sin\theta\sin(\varphi/2)/\sin\rho\right). \tag{A.2}$$

(The range of φ is taken to be $[-180°, 180°]$.) Then the coordinates (x, y) are

$$x = -\sqrt{2}\sin(\rho/2)\sin\gamma, \tag{A.3}$$

$$y = \pm\frac{1}{\sqrt{2}}\sin(\rho/2)\cos\gamma. \tag{A.4}$$

The sign of the equation for y is the sign of $\cos\theta$. Note that (x, y) lies within the ellipse $x^2 + 2y^2 = 1$. The equator $\theta = \pi/2$ lies along the major axis of the ellipse, and the North and South poles ($\theta = 0$ and $\theta = \pi$) lie at the points $(0, \frac{1}{2})$ and $(0, -\frac{1}{2})$ respectively. Lines of constant θ and φ are plotted in Figure A.1.

The inverse transformation is

$$\theta = \cos^{-1}\left(2y\sqrt{2 - x^2 - 4y^2}\right), \tag{A.5}$$

$$\varphi = -2\sin^{-1}\left(\frac{x\sqrt{2 - x^2 - 4y^2}}{\sin\theta}\right). \tag{A.6}$$

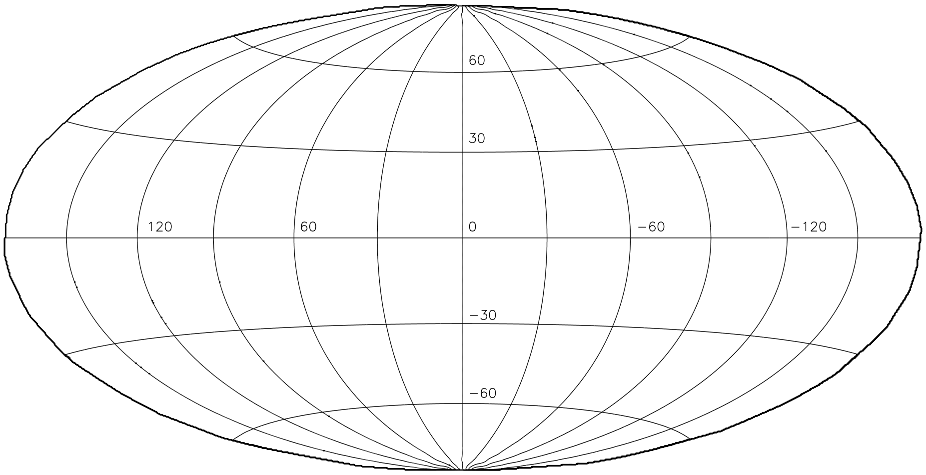

Figure A.1: The Aitoff Projection

In the unlikely event that there are any non-astronomers reading this, let us point out that Galactic longitude l is simply the azimuthal coordinate φ of a spherical coordinate system with its axis perpendicular to the Galactic plane and $l = 0$ in the direction of the Galactic center. Galactic longitude b is the complement of the spherical coordinate θ. Finally, let us point out that the transformation given here is appropriate for sky maps, in which the viewer is presumed to be situated inside the sphere. In Figure 2.4 the viewer is situated outside of the sphere, and so we have used the reflection of the above projection through the line $x = 0$.

Appendix B

Likelihoods and Weighted Averages

In this Appendix we prove the assertion made in Section 2.5 that the results of a likelihood analysis of the weighted-average map are identical to those of a joint likelihood analysis of the individual maps. We will prove the assertion for the case where two maps are averaged together, and then indicate how the generalization to several maps follows by induction.

We begin by establishing some notation. Let $\vec{x}_1$ and $\vec{x}_2$ be two independent measurements of the same underlying signal:

$$\vec{x}_i = \vec{s} + \vec{n}_i, \tag{B.1}$$

where the signal $\vec{s}$ and the noise $\vec{n}_i$ are Gaussian-distributed with zero mean and covariances

$$\langle \vec{s} \cdot \vec{s}^T \rangle = \mathbf{S}, \tag{B.2}$$

$$\langle \vec{n}_i \cdot \vec{n}_i^T \rangle = \mathbf{N}_i, \tag{B.3}$$

$$\langle \vec{n}_1 \cdot \vec{n}_2^T \rangle = \mathbf{0}, \tag{B.4}$$

$$\langle \vec{s} \cdot \vec{n}_i \rangle = \mathbf{0}. \tag{B.5}$$

In our case, the vectors $\vec{x}_i$ contain the $\mathcal{N}$ pixel values for each of two sky maps. The signal covariance matrix $\mathbf{S}$ is $\mathbf{Y} \cdot \bar{\mathbf{C}} \cdot \mathbf{Y}^T$. Let us assume further that the noise covariance matrices $\mathbf{N}_i$ are diagonal. The weighted average $\vec{y}$ of these two data vectors has components

$$y_j = \frac{x_{ij}/\sigma_{1j}^2 + x_{2j}/\sigma_{2j}^2}{1/\sigma_{1j}^2 + 1/\sigma_{2j}^2}, \tag{B.6}$$

where $\sigma_{ij}^2 = (\mathbf{N}_i)_{jj}$ is the variance of the noise in pixel j of data vector $\vec{x}_i$. Since the noise covariance matrices are presumed to be diagonal, we can write this expression in matrix form:

$$\vec{y} = (\mathbf{N}_1^{-1} + \mathbf{N}_2^{-1})^{-1} \cdot (\mathbf{N}_1^{-1} \cdot \vec{x}_1 + \mathbf{N}_2^{-1} \cdot \vec{x}_2). \tag{B.7}$$

We wish to compare the joint likelihood $L_1 = L(\vec{x}_1, \vec{x}_2)$ with the likelihood $L_2 = L(\vec{y})$ derived from the average map alone. Specifically, we wish to show that $L_1 \propto L_2$, where the proportionality constant does not depend on the signal covariance matrix $\mathbf{S}$.

Define a $2\mathcal{N}$-dimensional vector $\vec{X}$ to be the direct product of $\vec{x}_1$ and $\vec{x}_2$:

$$\vec{X} = \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \end{pmatrix}. \tag{B.8}$$

Then L_1 is the likelihood associated with $\vec{X}$. $\vec{X}$ is Gaussian-distributed with zero mean and covariance matrix

$$\langle \vec{X} \cdot \vec{X}^T \rangle = \begin{pmatrix} \mathbf{S} + \mathbf{N}_1 & \mathbf{S} \\ \mathbf{S} & \mathbf{S} + \mathbf{N}_2 \end{pmatrix}. \tag{B.9}$$

Likelihoods are unaffected (up to an uninteresting constant multiplicative factor) by a change of basis: we can replace $\vec{X}$ by $\vec{X}' = \mathbf{Q} \cdot \vec{X}$ for any nonsingular $\mathbf{Q}$ without changing L_1.[1] Let us make a change of basis so that

$$\vec{X}' = \begin{pmatrix} \vec{y} \\ \vec{\delta} \end{pmatrix}, \tag{B.10}$$

where $\vec{y}$ is given in equation (B.7), and $\vec{\delta} = \vec{x}_1 - \vec{x}_2$. Then the covariance matrix associated with $\vec{X}'$ is

$$M' = \begin{pmatrix} \langle \vec{y} \cdot \vec{y}^T \rangle & \langle \vec{y} \cdot \vec{\delta}^T \rangle \\ \langle \vec{\delta} \cdot \vec{y}^T \rangle & \langle \vec{\delta} \cdot \vec{\delta}^T \rangle \end{pmatrix}. \tag{B.11}$$

Let us examine the cross-covariance $\langle \vec{y} \cdot \vec{\delta}^T \rangle$:

$$\langle \vec{y} \cdot \vec{\delta}^T \rangle = (\mathbf{N}_1^{-1} + \mathbf{N}_2^{-1})^{-1} \cdot \langle (\mathbf{N}_1^{-1} \cdot \vec{x}_1 + \mathbf{N}_2^{-1} \cdot \vec{x}_2) \cdot (\vec{x}_1 - \vec{x}_2)^T \rangle. \tag{B.12}$$

Making the substitutions $\langle \vec{x}_i \cdot \vec{x}_i^T \rangle = \mathbf{S} + \mathbf{N}_i$ and $\langle \vec{x}_1 \cdot \vec{x}_2^T \rangle = \mathbf{S}$, we find that $\langle \vec{y} \cdot \vec{\delta}^T \rangle = \mathbf{0}$. Since there are no cross-correlations, the likelihood $L(\vec{X}')$ factors into $L(\vec{y})L(\vec{\delta})$. But $L(\vec{\delta})$ is a constant (independent of $\mathbf{S}$), so we have proved the result we were looking for:

$$L_1 = L(\vec{X}) \propto L(\vec{X}') \propto L(\vec{y}) = L_2. \tag{B.13}$$

---

[1] To see this, let $\mathbf{M} = \langle \vec{X} \cdot \vec{X}^T \rangle$ be the covariance matrix of $\vec{X}$. Then the covariance matrix of $\vec{X}'$ is $\mathbf{Q} \cdot \mathbf{M} \cdot \mathbf{Q}^T$. $L(\vec{X}') \propto \det^{-1/2}(\mathbf{Q} \cdot \mathbf{M} \cdot \mathbf{Q}^T) \exp\left(-\frac{1}{2}(\mathbf{Q} \cdot \vec{X})^T \cdot (\mathbf{Q} \cdot \mathbf{M} \cdot \mathbf{Q}^T)^{-1} \cdot (\mathbf{Q} \cdot \vec{X})\right) = L(\vec{X}) / \det \mathbf{Q}$.

The same result holds if more than two maps are averaged together. The easiest way to see this is by induction. Note that weighted averaging is associative: if we denote the operation of forming a weighted average by $\oplus$, then $\vec{x} \oplus \vec{y} \oplus \vec{z} = \vec{x} \oplus (\vec{y} \oplus \vec{z})$. We can therefore form a weighted average of arbitrarily many maps by repeatedly averaging maps together two at a time. None of these operations will change the likelihood.

Appendix C

Anisotropic Cosmological Models

In this dissertation we have generally restricted our attention to "standard" cosmological models based on the Friedmann-Robertson-Walker (FRW) solutions to the Einstein field equation. In particular, we have used the assumption that space is isotropic to argue that the angular power spectrum $C_l \equiv \langle a_{lm}^2 \rangle$ provides a complete description of the Gaussian statistics of the CMB anisotropy. There are certain cosmological models in which these assumptions are relaxed and this conclusion therefore does not follow. In these models, the covariance matrix

$$C_{lml'm'} \equiv \langle a_{lm} a_{l'm'} \rangle \tag{C.1}$$

need not be diagonal, and may depend on the azimuthal mode number m.

In this Appendix we will make brief comments on this class of anisotropic models. Specifically, we will consider models that are *locally* isotropic, and hence locally FRW, but that have nonstandard global topology. We will not discuss any of the other anisotropic cosmological models here. Among the most important of these are the Bianchi models, in which the assumption of isotropy is abandoned completely, while that of homogeneity is retained. CMB anisotropy spectra have been computed in simple cases for these models, which can be used to describe Universes in which there is large-scale shear or vorticity [7].

In Chapter 2 we pointed out that the usual assumptions of homogeneity and isotropy imply that spacetime must be locally isometric to one of the standard FRW solutions. Spacelike sections of these models look locally like portions of a three-sphere, Euclidean space, or hyperbolic space. The most natural assumption seems to be that these isometries are global rather than local. However, this need not be the case. All of the FRW models in principle admit solutions of nontrivial topology. Since the Einstein equation is local, these manifolds satisfy general relativity just as well as the standard FRW solutions.

The nontrivial solutions are easiest to visualize in the case of a flat Universe. For example, we can take a cube of side L and identify opposite sides to make a three-torus.[1] Furthermore, if we like we can replace the cube by an arbitrary rectangular solid with different lengths L_1, L_2, L_3 in the three directions. Although this is the simplest model to visualize, there are three other orientable topologies based on a rectangular solid as a fundamental cell, which are obtained by twisting some of the sides through 180° or 90° before making the identification. There are also two topologies based on a fundamental cell shaped like a hexagonal prism. A complete classification of these possibilities is given in [74]. If we like, we can let one or more of the length-scales in these models go to infinity (although if we let all three go to infinity, we're back to the standard topology).

Similar nontrivial topologies exist for the hyperbolic case, although they are, if possible, even uglier and less plausible than the flat toroidal models. There is, for example, the Lobell topology, which is based on a fundamental cell containing two hexagons and twelve pentagons [118]. There is one interesting difference between the flat and open compact geometries. In the flat case, the torus sizes in the three directions are completely arbitrary, while in the hyperbolic geometry the polyhedron's diameter is fixed with respect to the curvature radius. The Lobell topology is but one of an infinite number of possible choices for an open Universe [65, 165]. If the Universe is closed, on the other hand, the number of possible topologies is finite. For a complete classification, see [193].

The cosmological implications of nontrivial topologies have been explored for only the flat toroidal models.[2] If the torus sizes L_i are all much larger than the horizon distance, then the model is impossible to distinguish from one with the standard topology. If the torus size is smaller than the horizon, then there are a number of observational consequences, including changes in the statistical properties of the CMB anisotropy. Several people have analyzed the DMR data for signs of a finite torus size, concentrating on the cubical case $L_1 = L_2 = L_3$. As we shall see below, such a "small Universe" is expected to have less large-scale power than a Universe with the standard topology. The low DMR quadrupole therefore fits small-Universe models well. Jing & Fang [96] go so far as to claim a positive

---

[1] This spacetime is like that of the video game Asteroids, in which a spaceship that disappears off the right side of the screen reappears on the left.

[2] It comes as no surprise that no one has leapt to work out the CMB fluctuations in a Universe with the Lobell topology.

detection of such an infrared cutoff in the DMR data; other authors [50, 168] have performed more careful analyses and are more circumspect in their conclusions.

It is not hard to see why the topology of the Universe affects the CMB anisotropy spectrum. Consider the derivation of the Sachs-Wolfe angular power spectrum in Section 2.3. In equation (2.25) the angular power spectrum C_l is expressed as an integral over all wavevectors $\mathbf{k}$ in Fourier space:

$$C_l = \frac{1}{36\pi^4}\int d^3k P_\phi(k) j_l^2(kR) = \frac{1}{\pi^3}\int dk\, k^2 P_\phi(k) j_l^2(kR). \tag{C.2}$$

This equation simply expresses the fact that the CMB anisotropy is a superposition of all of the different Fourier modes of the gravitational potential. If the Universe is toroidal with some cell size L, then the Fourier transform of ϕ is discrete rather than continuous. This integral should therefore be replaced by a sum over all wavevectors $\mathbf{k} = (2\pi/L)(m_1, m_2, m_3)$, where $\{m_i\}$ are integers.

The dominant contribution to the integral in equation (2.25) comes from modes whose physical wavelength is approximately the angular scale of the anisotropy times the distance R to the last scattering surface. That is, most of the power comes from modes with $k \sim l/R$. The case that is of interest to us is when $L \sim R$,[3] and so for low l we expect the sum to be dominated by the first few modes. In other words, we expect the integral in equation (2.25) to be a good approximation for high l but not for low l. The DMR data are therefore a natural source of evidence for or against a toroidal Universe. It turns out that for $L \simeq R$ the quadrupole is strongly suppressed, and as L is reduced, the suppression is extended to the octupole and successive modes.

All attempts to compare the DMR data with the predictions of toroidal Universes have proceeded as if the usual rotationally-symmetric expression for the covariances of the a_{lm} were still correct. Specifically, the assumption is made that

$$\langle a_{lm} a_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'}. \tag{C.3}$$

However, since these models are not isotropic, this need not be the case. We can illustrate this with a direct computation. We begin by modifying equation (2.22), which expresses the Sachs-Wolfe contribution to a_{lm} in terms of the Fourier transform of the gravitational potential, to correspond to the toroidal topology. This involves simply replacing the integral

---

[3] If $L \ll R$, then we would see the periodicity of space directly, while if $L \gg R$ the torus size is much larger than the horizon, and there is no hope of seeing any effect at all.

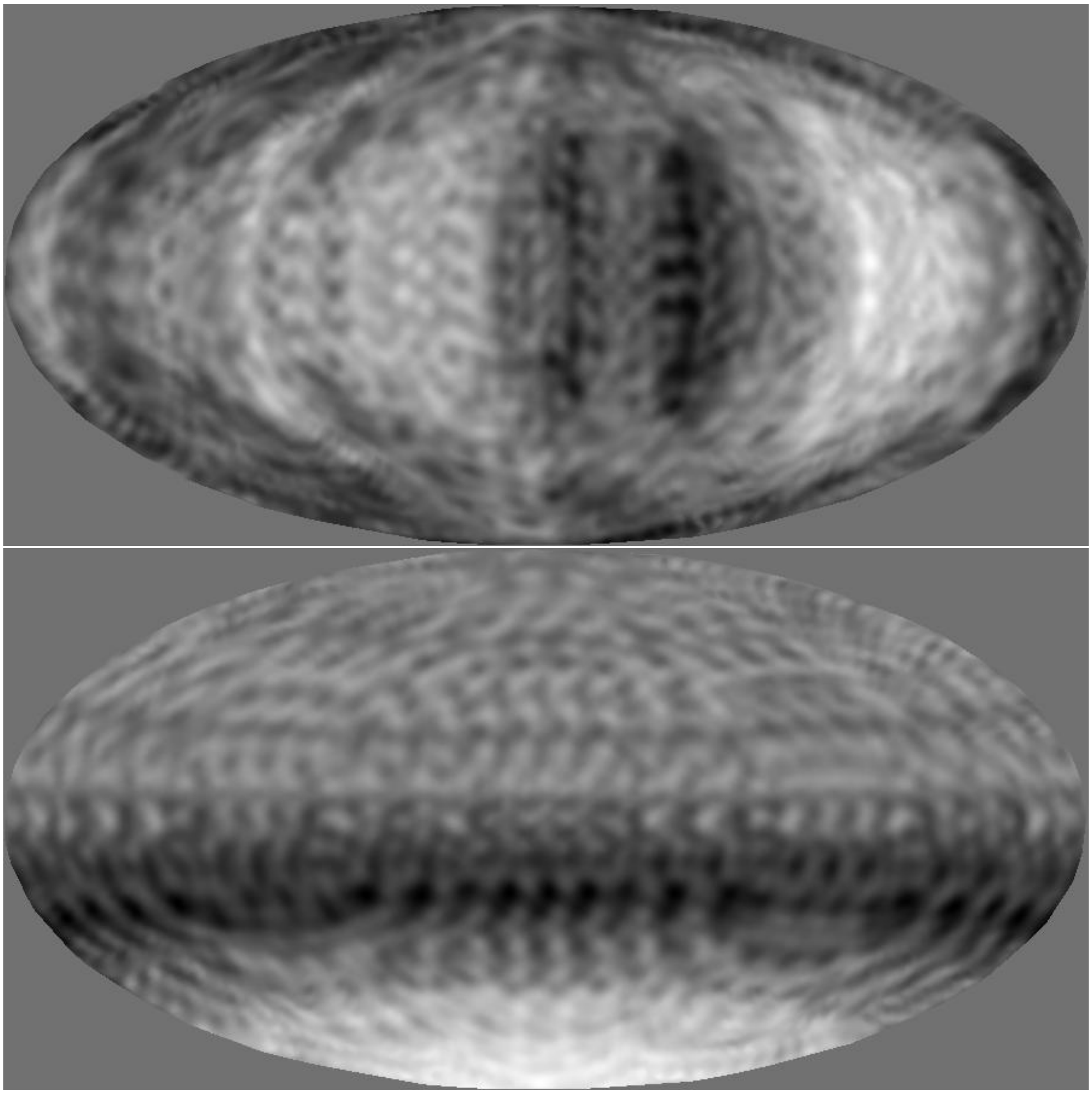

Figure C.1: CMB anisotropy in a toroidal Universe

We show simulated sky maps of the CMB anisotropy in a toroidal Universe with two different choices of cell sizes. The top panel shows a model in which the $L_1 = L_2 = 5R$ and $L_3 = 0.2R$, while the bottom is a model in which $L_1 = L_2 = 0.2R$ and $L_3 = 5R$. Both were made assuming pure Sachs-Wolfe fluctuations with an $n = 1$ power spectrum.

over $\mathbf{k}$ by a discrete sum:

$$a_{lm} \propto \int d\Omega \sum_{\mathbf{k}} Y_{lm}(\hat{\mathbf{r}}) e^{-iR\mathbf{k}\cdot\hat{\mathbf{r}}} \tilde{\phi}_{\mathbf{k}}. \tag{C.4}$$

We then form the product $a_{lm}a_{l'm'}$ and take an ensemble average:

$$\langle a_{lm}a_{l'm'}\rangle \propto \int d\Omega \int d\Omega' \sum_{\mathbf{k}} Y_{lm}(\hat{\mathbf{r}}_1) Y_{l'm'}(\hat{\mathbf{r}}_2)\, P_\phi(k)\, e^{iR\mathbf{k}\cdot(\hat{\mathbf{r}}_2-\hat{\mathbf{r}}_1)}. \tag{C.5}$$

This equation is analogous to equation (2.23). We can expand the exponential in spherical harmonics as in Chapter 2, to find that

$$\langle a_{lm}a_{l'm'}\rangle \propto \sum_{\mathbf{k}} j_l(kR) j_{l'}(kR) Y_{lm}(\hat{\mathbf{k}}) Y_{l'm'}(\hat{\mathbf{k}}). \tag{C.6}$$

This sum does not vanish in general when $(l,m) \neq (l',m')$. In the limit $L \to \infty$, the sum over $\mathbf{k}$ becomes an integral, and the off-diagonal correlations do tend to zero because of the orthogonality of the spherical harmonics.

In the case of a cubic toroidal model, this sum turns out to be quite small when $(l,m) \neq (l',m')$. This is perhaps not surprising, since space is in some heuristic sense "nearly isotropic" in these models. Previous analyses of these models are therefore probably not significantly compromised by their neglect of off-diagonal correlations. However, if one wishes to consider the more strongly anisotropic models in which the three fundamental length-scales of the torus differ, off-diagonal elements are quite important. One way to see this is to make simulated sky maps based on such toroidal models. We show such maps in Figure C.1 for two interesting cases. It is clear that the structures in the maps have characteristics that are not captured by a simple isotropic power spectrum, and that a simple power-spectrum analysis may not be sufficient to test these models.