Cosmological Density and Power Spectrum from Peculiar Velocities: Principal Component Analysis and Nonlinear Effects

Lior Silberman

March 26, 2000

Contents

Abstract

Likelihood analysis has emerged as a central tool in recovering cosmological information from observational data. In particular, it allows recovering the power spectrum even in the face of inhomogeneous data, such as the current peculiar velocity catalogues. However, the prior models used for the analysis are incomplete with respect to the handling of nonlinear phenomena. Also, the large-scale linear regime data are very noisy. Analyzing peculiar velocity catalgoues, we explore ways to overcome these difficulties by using Principal Component Analysis to extract more information on the relation between the model, the data and the analysis results, and to understand the significant part of the data. On the other hand, by 'breaking' the power spectrum we improve the parameter estimation by using the strong areas of the model in the analysis. This is tested on mock catalogues. Applying both techniques we extend previous results and reach a lower value for the density parameter, $\Omega \approx 0.35$.

This work was done for the Honors Program senior thesis requirement under the direction of Prof. Avishai Dekel, the Racah Institute of Physics, the Hebrew University of Jerusalem, Israel.

1 Theoretical Background

In the standard Big Bang model, structure in the universe forms as small density perturbations generated in the early universe are amplified by gravitation. This process, known as *Gravitational Instability,* results from the fact that (relatively) denser regions attract nearby mass, which appears to be 'repelled' by the sparser regions. The initial fluctuations are thought to be quantum fluctuations in the early universe, themselves amplified by inflation.

The statistics of the fluctuations is usually modeled as a Gaussian Random Field. In this case the amplitude of the Fourier transform of the fluctuations, namely the *Power Spectrum,* determines the entire statistics. It depends on various cosmological parameters, and it is therefore natural to attempt to reconstruct the original power spectrum from the present-day state of the universe in order to estimate these parameters. After an investigation of the relation between the initial and present states, i.e. the evolution in time, follow brief discussions of theoretical models for the power spectrum, Gaussian random fields, and finally the data used in this research, both real-universe observations and mock catalogues prepared from numerical N-body simulations.

1.1 Density and Velocity fields, Gravitational Instability, Fluctuations

When the universe has cooled enough, and the density is sufficiently low, one can ignore the non-gravitational interaction of the mass in the universe. The dynamical parameters of the mass distribution are the mass density, and the local velocity. For large enough scales, we assume the mass to have a uniform density and to flow according to Hubble's law (homogeneity and isotropy), so we need only consider the perturbation from this average behaviour. Following the development in [10], density is written in terms of the *density contrast,* $\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}}$, and the velocity in terms of the *peculiar velocity* (i.e. relative to the Hubble velocity). In comoving coordinates (in the non-relativistic approximation) the equations of motion for the perturbation fields are:

$$\begin{cases} \triangle\Phi = \frac{3}{2}H^2\Omega\delta & [poisson's\ equation] \\ \dot{\delta} + \vec{\nabla}((1+\delta)\vec{u}) = 0 & [continuity] \\ \dot{\vec{u}} + H\vec{u} + (\vec{u}\vec{\nabla})\vec{u} = -\vec{\nabla}\Phi & [Euler's\ equation] \end{cases} \tag{1}$$

where distance is measured in units of velocity (i.e. as Hx, H being the current Hubble constant). Taking the linear approximation, valid for small perturbations (i.e. for small δ and $\vec{u}$), and separating the time and space coordinates, we get for the time component an equation of the form

$$\ddot{\delta} + 2H\dot{\delta} = \frac{3}{2}H^2\Omega\delta \tag{2}$$

In a particular cosmology (which determines $H(t);\Omega(t)$) there are two solutions for $\delta(t)$, one growing and one decaying. In this simple model, fluctuations on different scales grow at the same rate[1], usually written as $D(t)$ (neglecting the decaying mode). This phenomenon is known as *Gravitational Instability*, since the fluctuations tend to grow with time. In terms of the growth rate, the equation of continuity is: $\delta = -\frac{HD}{\dot{D}}\vec{\nabla}\vec{u}$. Defining $f = \frac{d\ln D}{d\ln a} = \frac{\dot{D}}{HD}$, we have the *velocity-density* relation:

$$\delta = -\frac{1}{f}\vec{\nabla}\vec{u} \tag{3}$$

The parameter f depends on the cosmological parameters in a complicated way. Lahav et. al. in [8] give an approximation :

$$f(\Omega,\Lambda) \approx \Omega^{0.6} + \frac{\Lambda}{70}(1+\frac{\Omega}{2}) \tag{4}$$

The values considered here for the cosmological constant are small enough to safely assume $f(\Omega) \approx \Omega^{0.6}$.

The velocity-density relation has immediate implications in the Fourier domain. Since the growth rate in only a function of time, the Fourier transforms of the density and velocity are related as:

$$\tilde{\delta}(\vec{k}) = \frac{1}{f}(i\vec{k})\tilde{\vec{u}}(\vec{k}) \tag{5}$$

[1] In the presence of massive neutrinos (HDM), small-scale fluctuations are suppressed due to free-streaming. This makes the growth rate a function of scale. See, e.g. [2]

Assuming isotropy, these are actually functions of the modulus k only. In the discussion that follows, most of the attention centers on the *Power Spectra*, $P_\delta(k) = \left\langle |\tilde{\delta}(k)|^2 \right\rangle; P_v(k) = \langle |\tilde{v}(k)|^2 \rangle$ and we see immediately that:

$$P_\delta(k) = \frac{1}{f^2(\Omega)} k^2 P_v(k) \tag{6}$$

Thus, in the linear approximation the velocity and density power spectra can be used interchangeably. Also, since the fluctuations grow in time proportionally to $D(t)$, the power spectra also evolve in time proportionally (according to $D^2(t)$). Thus, in the linear regime, the present-day and initial power spectra are related by a simple constant. This separates the discussion of the power spectrum into two parts, the *shape* and the *normalization.*

1.2 Power Spectra, COBE-normalized CDM

Having found a simple model for the evolution of the mass density and velocity perturbations with time, we must now turn our attention to the question of the variation in space, which is essentially the question of the initial conditions. The Power Spectrum of the initial fluctuations depends on the cosmological model. It is affected by the cosmological parameters (mass density, cosmological constant, Hubble expansion factor), as well as the nature of the matter in the universe and the inflationary model (e.g. the fraction of mass present as baryonic matter, as neutrinos, etc). It is also affected by the so-called primordial power spectrum, generally taken to have the shape $P(k) \propto k^n$. As long as the non-gravitational interaction in the universe is strong enough, the power spectrum of these primordial fluctuations changes in time, while various effects come into play (e.g. free-streaming of relativistic particles). As the universe cools and the density drops, these interactions have a smaller effect, and the theory of the previous section shows that the power spectrum stops evolving. This is not strictly true, since in later times fluctuations collapse sufficiently to leave the linear regime. The 'broken' power spectrum model is an attempt to handle this difficulty.

Most of this work was done in the framework of the CDM model, for which the (linear) power spectrum can be written as:

$$P(k) = AT^2(\Omega, \Omega_b, h; k) k^n \tag{7}$$

where A is a normalization constant, and T is the *transfer function.* Continuing the work of Zaroubi et. al. [18] and Freudling et. al. [3] we use the transfer function and normalization calculated by Sugiyama in [11]:

$$\begin{gathered} T(k) = \tfrac{\ln(1+2.34q)}{2.34q} \left[1 + 3.89q + (16.1q)^2 + (5.46q)^3 + (6.71q)^4\right]^{-1/4} \\ q = k\left[\Omega h \exp(-\Omega_b - (2h)^{1/2}\Omega_b/\Omega)\right]^{-1} \end{gathered} \tag{8}$$

Ω_b is the mass density of baryons in the universe taken as $0.024h^{-2}$, the value used by [18] ($H = h \cdot 100km \cdot s^{-1}Mpc^{-1}$).

The normalization factor A is calculated by relating the anisotropies in the Cosmic Microwave Background to the density fluctuations at the time of recombination, using the COBE satellite DMR experiment to measure the power spectrum at extremely large scales. In fact, with more accurate measurements of the CMB anisotropies it should be possible to directly measure the power spectrum and fit it with a theoretical model. Since the actual distance scale related to a (directly measured) angular scale depends on the curvature of the universe (i.e. the cosmological model), this normalization also depends on the cosmological parameters, since the linear scale corresponding to the angular measurements changes. We quote here the normalization in case of the (flat) ΛCDM model without tensor fluctuations used in this work: $A_{COBE} = \exp(a_1(n))\exp(a_2(\Omega))$

$$\begin{gathered} a_1(n) = 2.78(n-1) \\ a_2(\Omega) = 7.83 - 8.33\Omega + 21.31\Omega^2 - 29.67\Omega^3 + 10.65\Omega^4 + 15.42\Omega^5 - \\ 6.04\Omega^6 - 13.97\Omega^7 + 8.61\Omega^8 \end{gathered} \tag{9}$$

Another form of normalization from smaller scales is to measure some statistic of the random field determined by the power spectrum, e.g. the mean fluctuation in a region of a certain size. Given a window function $W(x)$, the mean fluctuation is:

$$\sigma_r^2 = \left\langle \int \delta^2(x - x')W(x'/r)d^3x' \right\rangle = \int P_\delta(k)\tilde{W}^2(kr)d^3k \tag{10}$$

The usual case is the top-hat window, $W(x) = \frac{3}{4\pi r^3}\theta(1 - |x|)$ (θ is the Heavyside function), where the resulting statistic is written as σ_r. Statistics of cluster abundance indicate that for the $8h^{-1}Mpc$ scale, $\sigma_8\Omega^{0.56} \approx 0.57 \pm 0.05$ ([14]), so this can be used to normalize the power spectrum. When using a different normalization (COBE, data-determined, etc), this value can be used to test the fits to the power spectrum.

1.3 Gaussian Random Fields

A Gaussian Random Field is defined as a random field $F(x)$ such that for every n points in the field, the joint probability distribution of the field values at these points is a multivariate Gaussian, i.e. the probability density is: ($f_i = F(x_i)$)

$$P(f_1, \ldots, f_n) = [(2\pi)^n \det \mathbf{C}]^{-1/2} \exp^{-\frac{\chi^2}{2}}; \quad \chi^2 = -\frac{1}{2}\mathbf{f}^\dagger\mathbf{C}^{-1}\mathbf{f} \tag{11}$$

An immediate calculation gives:

$$\langle f_i f_j \rangle = C_{ij} \;\Rightarrow\; C = \langle \mathbf{f}\mathbf{f}^\dagger \rangle \tag{12}$$

Defining the *two-point correlation function* as:

$$\xi_F(x, x') = \langle F(x)F(x')\rangle \tag{13}$$

and assuming homogeneity and isotropy, $\xi_F(x, x')$ is only a function of the distance between x, x', and:

$$C_{ij} = \xi_F(|x_i - x_j|) \tag{14}$$

The statistics is then governed by the correlation function. Calculating the Fourier transform of the correlation function gives:

$$\tilde{\xi}_F(k, k') = \left\langle \tilde{F}(k)\tilde{F}(k')\right\rangle = P_F(k)\delta(k - k') \tag{15}$$

Where $P_F(k)$ is the power spectrum and δ is the Dirac delta function. This relates the power spectrum of the field to the correlation function. We will now make the assumption that the initial fluctuations of the density and velocity fields are Gaussian random fields. Since the time evolution of the fields is simply multiplication by an overall factor $D(t)$, these fields will remain Gaussian in the linear regime. As the perturbations grow and nonlinear evolution starts, the fields cease to obey this approximation, and one must either use a more appropriate probability distribution, or find ways to minimize the effects of this on the analysis.

Specializing to the peculiar velocities, we actually have a field with three components, and with a correlation tensor

$$\xi_{\mu\nu}(x - x') = \langle u_\mu(x)u_\nu(x')\rangle \tag{16}$$

We note that the velocity correlation tensor depends on the direction of the vector $r = x - x'$ with respect to the selected system of coordinates. In linear theory we can write:

$$\xi_{\mu\nu}(\mathbf{r}) = \xi_\perp(|r|)\delta_{\mu\nu} + \left[\xi_\parallel(|r|) - \xi_\perp(|r|)\right]\hat{r}_\mu\hat{r}_\nu \tag{17}$$

From Earth the measurable component of an object's velocity is the radial component (via the redshift), the statistics of which is also Gaussian. In terms of the above functions the correlation function is:

$$\xi_u(x, x') = \hat{x}^\mu\xi_{\mu\nu}(x - x')\hat{x}'^\nu = \xi_\perp(|r|)cos(x, x') - \left[\xi_\parallel(|r|) - \xi_\perp(|r|)\right]cos(x, r)cos(x', r) \tag{18}$$

Where the radial correlation functions are given by (using the linear relation between the power spectra, equation 6):

$$\begin{gathered}\xi_{\perp,\parallel}(r) = \tfrac{f^2(\Omega)}{2\pi^2}\int_0^\infty P_\delta(k)K_{\perp,\parallel}(kr)dk \\ K_\perp(x) = j_1(x)/x\ ;\ K_\parallel(x) = j_0(x) - 2j_1(x)/x\end{gathered} \tag{19}$$

With $j_l(x)$ the spherical Bessel function of order l. Note that distance is here measured in redshift units. In metric units there is an additional factor of H^2.

To summarize, for a given power spectrum and a set of locations the probability distribution of the velocity field at these locations is determined by the correlation matrix $S_{ij} = \xi_u(x_i, x_j)$.

Regarding measurement errors, we assume at each point x_i an independent random error e_i, distributed as a Gaussian with zero mean and variance n_i. The error vector is a multivariate Gaussian with a diagonal correlation matrix N, such that $N_{ij} = \delta_{ij} n_i$. Then the probability distribution of the *observed* peculiar velocities will be the distribution of the sum $u_i + e_i$. The sum of multivariate Gaussians is Gaussian, with the correlation matrix being the sum of the correlation matrices. That is, the correlation matrix of the observed velocities will be $C = S + N$.

1.4 Real-Universe Data and Mock Catalogues

In the last section, the discussion focused on the statistics of the radial velocity field, rather than the density field. The density field is difficult to observe directly, since almost all of the mater of the universe is dark. The galaxies (which are the objects amenable to observation in the relevant distance scales) represent only a small part of the total mass in the universe. Also, the distribution of the galaxies does not follow directly the distribution of the mass. For example, galaxies tend to form in regions of high density. Thus using the galaxy number distribution leads to a *bias* which is difficult to measure. Since we have seen that in the linear approximation the power spectrum of the velocity distribution is (essentially) the same as that of the mass density, it is reasonable to try to recover the power spectrum of the velocity field. Assuming that the peculiar velocity of observed galaxies reflects that of the surrounding dark matter, this would give a more accurate estimate for the power spectrum. The possible *velocity bias* which would result if this approximation was invalid was neglected here.

Thus, the goal is to recover the power spectrum of the density field, using a discrete sample of the peculiar velocity field obtained by measuring the peculiar velocities of galaxies. We use catalogues which contain for each galaxy the direction to the galaxy, the redshift, and an estimate for the distance. The measurement of distances over cosmological distance scales entails a significant error, which is the greatest source of observational error in the following analysis.

The observational data used were two peculiar velocity catalogues, the Mark III catalogue ([15],[16],[17]) and the SFI catalogue ([4]). In the Mark III catalogue galaxies were grouped to reduce nonlinear effects while lowering the distance measurement error. In the SFI catalogue, in order to avoid selection effects caused by the inclusion of nearby small linewidth galaxies, only galaxies with linewidth $\log w \geq 2.25$ were included in the analysis. The catalogues were also corrected for Malmquist bias.

In order to evaluate various reconstruction methods, it is necessary to have a sample for which the true state of affairs is known. In this case, this means simulated catalogues drawn from a 'universe' in which the power spectrum is known. We also need these mock catalogues to reflect the various distortions inherent in the real-universe observation process, such as measurement errors, selection effects, various biases, etc. Several versions of mock catalogues were used.

Initially, mock catalogues were generated by generating random perturbations to points uniformly distributed on a grid according to a predetermined power spectrum. The catalogues thus obtained contained no errors. Random errors were added later, but to the velocity rather than to the distances. This created an 'idealized' catalogue. With this kind of mock catalogue it is possible to isolate specific properties of the method of analysis, e.g. the effects of an incorrect error model or of using a Gaussian error model for non-Gaussian errors, without other uncertainties or biases.

In addition to these simplified catalogues, two 'realistic' sets of mock catalogues were used. The first set, described in [7], was generated by a simulation whose initial conditions were based on a reconstruction of the smoothed density field of the IRAS 1.2Jy redshift survey, evolved back in time, with small-scale power added. The system was then evolved forward in time using an N-body simulation, assuming $\Omega = 1$.

The second set is based on the recent unconstrained GIF simulation ([6]), with an initial power spectrum of ΛCDM with $\Omega = 0.3, \Lambda = 0.7, h = 0.7$. The latter simulation employed a finer grid spacing and smaller particles, thus allowing higher resolution. The catalogues were in several versions with different cluster identification radii, to vary the degree of smoothing of nonlinear effects in the catalogue. In both cases the mock catalogues attempt to reproduce the selection effects of the real-universe catalogues, as well as any known biases and observational errors present there.

1.5 Likelihood Analysis

Given the velocity catalogue, in other words a sample withdrawn from a Gaussian random field, we would like to determine the power spectrum and the cosmological parameters. We require that the reconstruction includes an estimate of the errors in the values of the parameters. The direct approach based on calculating the discrete Fourier transform of the data to get the power spectrum is problematic due to the noisy data. We therefore prefer an indirect method.

In formal terms, each model is a theoretical probability distribution, $\mathcal{P}(\mathbf{d}|\mathbf{m})$ - the probability of the data given that model. For the inverse problem - finding the best model given the data, we use Bayes' theorem:

$$\mathcal{P}(\mathbf{m}|\mathbf{d}) = \frac{\mathcal{P}(\mathbf{m})\mathcal{P}(\mathbf{d}|\mathbf{m})}{\mathcal{P}(\mathbf{d})} \tag{20}$$

Where we assume that $\mathcal{P}(\mathbf{m})$ and $\mathcal{P}(\mathbf{d})$ are the a-priori probabilities for the model and data, respectively. This is the *Likelihood* function, $\mathcal{L}(\mathbf{m}|\mathbf{d})$, which in the Bayesian view gives us the likelihood that the model is correct given the data. As long as we only have one set of data, $\mathcal{P}(\mathbf{d})$ is merely a normalization constant. Lacking a way to evaluating this parameter, we cannot use this method to compare different models for different data. Also, without a better alternative, we will assume $\mathcal{P}(\mathbf{m})$ to be a constant over the range of models we consider. We are therefore taking $\mathcal{L}(\mathbf{m}|\mathbf{d})$ to be proportional to $\mathcal{P}(\mathbf{d}|\mathbf{m})$. Under the Bayesian view of probability, the concept of the likelihood of the model given the data is meaningful even when we do not have a frequentist probability for the model.

The Likelihood analysis reconstruction method fits to the data that model which maximizes $\mathcal{L}$. This Bayesian approach can be justified by a frequentist theorem, which states that in the limit of large data sets, the maximum-likelihood estimate is the best estimate for the parameters[2].

In the present case, for the data vector d, for each set of cosmological parameters Θ we have a correlation matrix $C(\Theta)$, giving a likelihood function (see equation 11):

$$\mathcal{L}(\Theta) = \left[(2\pi)^n \det \mathbf{C}(\Theta)\right]^{-1/2} \exp^{-\frac{d^\dagger C(\Theta)^{-1} d}{2}} \tag{21}$$

It is more convenient to work with the logarithm of this function. Dropping the additive constant $-n\ln(2\pi)$ we also use (somewhat inconsistently):

$$\mathcal{L}(\Theta) = -\frac{1}{2}\left(ln \det C + d^\dagger C^{-1} d\right) \tag{22}$$

We can now scan the parameter space, or use better techniques for nonlinear optimization, to find the best-fit parameters. We estimate the confidence levels on the parameters by observing the sharpness of the extremum. The 1σ confidence limit is taken at the point where of a drop of 1 in the log-likelihood. A more refined approach (see [12] for details) uses the second derivatives of the likelihood function, i.e. the Fisher matrix:

$$F_{ij} = \left\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_i \partial\theta_j} \right\rangle \tag{23}$$

The matrix F^{-1} can then be used as an estimate of the covariance of the *parameters* (when $\mathcal{L}$ is used as a probability distribution in parameter space).

The confidence level on the parameters only depends on the *relative* likelihood at different points in parameter space. We also need an absolute indication for the goodness-of-fit of the method. The most commonly used estimate the χ^2 statistic. Defining the data matrix $D = dd^\dagger$ we write:

$$\chi^2 = d^\dagger C^{-1} d = Tr\left[C^{-1} D\right] \tag{24}$$

Since C is the ensemble-average of D, we expect $C^{-1}D$ to be close to the unit matrix for a large enough survey, or χ^2 to be close to the number of degrees of freedom. Assuming this

[2]The result is presented in [12] which include further references

statistic to have a χ^2-distribution with n degrees of freedom, this means $\frac{\chi^2}{m} \approx 1 \pm 2\sqrt{\frac{2}{m}}$ within the 2σ confidence limit.

1.6 Likelihood analysis of the Mark III and SFI catalogues

The Likelihood analysis technique has been used to recover the cosmological parameters by Zaroubi et. al. ([18]) for the Mark III catalogue, and by Freudling et. al. ([3]) for the SFI catalogue. The method was validated using the mock catalogues of [7]. For the $n = 1$ ΛCDM model considered here the both the Mark III and SFI studies found a degeneracy with respect to Ω and h. For Mark III the maximum-likelihood ridge was approximately: $\Omega h_{50}^{1.3} = 0.83 \pm 0.09$. and for SFI: $\Omega h_{60}^{1.3} = 0.59 \pm 0.07$. To simplify the analysis and break this degeneracy we assume throughout $h = 0.65$ (found in independent studies of supernovae type Ia).

For this model, ΛCDM with $n = 1$; $h = 0.65$, the best-fit value for the density parameter for Mark III was: $\Omega \approx 0.56 \pm 0.04$, while for SFI it was: $\Omega \approx 0.51 \pm 0.05$. The χ^2 values in both cases are very close to unity. From the power spectra at these parameter values, we get $\sigma_8\Omega^{0.56} \sim 0.91$ for Mark III, $\sigma_8\Omega^{0.56} \sim 0.82$ for SFI, with about 10% error margins. Both the values for the density parameter and for the $\sigma_8\Omega^{0.56}$ are higher than values commonly reported (e.g. in supernovae type Ia or redshift surveys studies). In particular, the values for $\sigma_8\Omega^{0.56}$ are higher than the cluster-normalized value quoted above.

In these studies, several tests were made to ensure the robustness of the results. Testing with the mock catalogues showed that the method recovers the simulation power spectrum with small systematic errors. It must be bourn in mind that these simulations had a limited resolution, and thus artificially weak nonlinear effects. This may not be the case for the real universe, and a part of the current study was an attempt to handle this. Regarding the error estimates, various modifications (with free parameters) to the error model were tested using the likelihood analysis mechanism itself. The results were essentially in agreement with the original error estimates.

The most significant problem with the analysis is the fact that partition of the data leads to conflicting results. In both catalgoues, separating the galaxies by the median radial distance into two data sets and analyzing each half separately gives different results. The nearby data prefer a higher power spectrum (in parameter space, a higher value of Ω) than the distant half. This effect was not observed in the simulation. This may be due to a shortcoming in the power spectra used in the analysis, as some results here seem to indicate, but may also be due to difficulties in the error model. With a $\sim 15\%$ distance error for individual galaxies, and an average absolute radial velocity of about $550 Km/s$ (SFI), the median distance ($46h^{-1}Mpc$) is also the distance in which the errors are on the order of magnitude of the signal. In the second part of this work we develop the PCA technique which sheds some light on this problem, and use the 'broken' power spectrum scheme to improve on the analysis.

2 Principal Component Analysis

As discussed earlier, the method of likelihood analysis is theoretically optimal. In reality, however, we face several problems. Even if the data is correctly described by linear theory, it is still quite noisy - the error of the distance measurements may be larger than the peculiar velocities! The grouping of some of the galaxies improves the situation, but not for the whole sample. Also, it is not at all clear that the errors are indeed Gaussian, or that the error estimates are accurate. With regard to the model, also, the real data contains nonlinear phenomena. Moreover we can never be sure that the power-spectrum of the real universe is well approximated by one of the models under consideration, i.e. some selection of the model parameters.

Thus it is advantageous to find a way to contrast the model and the data using the 'strong points' of the model as it were, while minimizing the effect of the areas in our model which are less reliable. In this context, the 'model' means the entire description of the statistics of the data, which is what our method of analysis attempts to do. The cosmological model, but also the individual error estimates and the distribution chosen are all part of the 'machinery'. Likelihood analysis, under the assumption of Gaussianity, is essentially the recovery of a 'meta-parameter' - the correlation matrix. The method as described earlier attempts to fit correlation matrices to the measured data. If it were possible to restrict the method to only use the parts of the correlation matrix we are most sure of, we might expect to fare better.

This simplistic view is not practical, as this information is contained in all the components of the matrix. For example, when we consider both the linear and nonlinear description of the power spectrum, in the large scales the data is probably linear, but in the small scales probably not, and each component of the correlation matrix is an integral over the entire power spectrum, with different components having different contributions from different scales. We introduce here new tools to analyze the data, possibly including also a reduction of the data. The next section considers adjusting the model so that the fit for the interesting parameters is governed by the parts of the data which are best described by the model.

2.1 Change of variables

Considering the problem from the viewpoint of the data, we wish to extract from the large, noisy data vector as much information as possible, in the ranges best described by our model. If this also reduces the amount of the information actually used we will also have computational benefits. Calculating the likelihood function defined above entails two significant calculations: First, the integration of the power spectrum with respect to the two kernels $K_{\perp,\parallel}$. This must be (in theory) done for each possible separation of two galaxies in the survey. Then, one has to invert the resulting correlation matrix. The catalogues used here have about 10^3 galaxies, so these are formidable tasks. We would like to extend the analysis to include future redshift surveys such as 2dF and SDSS which will contain $10^5 - 10^6$ galaxies, when calculating the full likelihood function will be impractical. The cost of the integrations can be mitigated by tabulating the integrals for a set of distances and then interpolating the results, but even for the current catalogues these tasks take about 2-3 seconds on the workstation used for each point in parameter space. Reducing the size of the correlation matrix shortens the inversion time significantly.

In the attempt to reduce the data, simply ignoring the noisy (distant) data and keep only the nearby part will not do. While the potential flow approximations shows that the velocities are affected by the density field from large distances, using data from a small area will still reduce the range of scales probed by the analysis. Also, it is necessary to use the more distant data since the smaller scales are the ones most prone to nonlinear effects. We also note that if the Gaussian error model is correct, then incorporating the noisy data into the analysis should not introduce any bias to the result, and should even reduce the error bars. This is because the analysis takes into account the error inherent in the measurements. If the error or data models are not accurate, we could benefit from a more sophisticated reduction of the data, which will include some information from larger scales.

In many cases, a different representation of the same data can lead to new insights. Vogeley & Szalay in [13] suggested using Principal Component Analysis, or PCA, to find an optimal representation of the data, in terms of which the analysis is easier. While simply using a new representation of the same data will not change the results, combining this new representation with the idea of truncation, i.e. of using only part of the data, may improve the results. Working the other way around, as in the note in the next section, based on [12], shows that PCA is the optimal way to truncate the data.

Recalling the discussion of likelihood analysis, the statistics of the (n-dimensional) data vector d, is described by the correlation matrix $C = \left\langle dd^{\dagger} \right\rangle$. We seek to reexamine the data in terms of new variables in a way that would best represent the information in the survey.

We will restrict our attention to linear transformations since then if we define a new set of m variables by $\tilde{d} = \Psi d$, Ψ being an $m \times n$ matrix, the statistics of the new variables is still that of a multivariate Gaussian governed by the correlation matrix

$$\widetilde{C} = \left\langle \widetilde{d}\widetilde{d}^{\dagger} \right\rangle = \Psi C \Psi^{\dagger} \tag{25}$$

We immediately see that if we keep all the original information, i.e. have Ψ invertible, we have done nothing, since then $Tr\left[\tilde{C}^{-1}\tilde{D}\right] = Tr\left[C^{-1}D\right]$. Having said this, we must now choose the new set of variables. The first idea is to have the new variables uncorrelated. Uncorrelated variables are better for our purpose since we will only use a part of the information, since correlations between the data included in the likelihood analysis and the data left out might lead to systematic errors.

This is equivalent to taking $\widetilde{C}$ to be diagonal. Since the original correlation matrix is symmetric by definition, it is possible to diagonalize it with a transformation like the one above, with the rows of the matrix Ψ corresponding to the eigenvectors of the matrix C. These eigenvectors are called the *Principal Components* of the correlation matrix.

We will now use this construction in two different ways. First, we can redo the likelihood analysis in terms of the new variables. Using all the variables will just repeat the previous analysis, so we should first select a subset with which to work. This subset should (generally) be the modes with the greatest eigenvalues, as shown later. Secondly, in addition to the use of this transformation in the likelihood analysis, we might learn something about our model and survey by examining both the eigenmodes and the original analysis on a mode-by-mode basis.

Regarding the choice of correlation matrix: since in actual implementations we do not know the correlation matrix of the data (that is essentially what we are trying to determine!), this analysis required a prior - some predetermined correlation matrix C^0. This matrix is conventionally taken to be the best-fit correlation matrix from the full likelihood analysis. For the present discussion this is sufficient, as we only use PCA to analyze the results of this likelihood analysis, rather than perform likelihood analysis using the transformed data $\tilde{d}$. If it appears that using one prior is not sufficient, an iterative approach is possible, where the best-fit result of one iteration is used as the prior for the next.

We plot here the results for the Mark III catalogue and the ΛCDM power spectrum, with the prior taken from the original (no-PCA) results: $\Omega = 0.55$; $h = 0.65$; $n = 1$. Figure 1 shows the eigenvalues for the correlation matrices we use. The 'whitened matrix' is described in the next section.

A Note: General Transformations

A general formalism, discussed by Tegmark et. al in [12], shows how to choose the proper transformation. We write the likelihood function as:

$$\mathcal{L}(\Theta) = \frac{1}{2} Tr \left[\ln C(\Theta) + C(\Theta)^{-1} D \right] \tag{26}$$

Defining $C_{,i} = \frac{\partial C}{\partial \theta_i}$ and $A_i = C^{-1} C_{,i} = (\ln C)_{,i}$, we write the derivative as:

$$\mathcal{L}_{,i} = \frac{1}{2} Tr \left[C^{-1} C_{,i} - C^{-1} C_{,i} C^{-1} D \right] \tag{27}$$

In particular we note that since $\langle D \rangle = C$, we have $\langle \mathcal{L}_{,i} \rangle = \frac{1}{2} Tr \left[C^{-1} C_{,i} - C^{-1} C_{,i} C^{-1} C \right] = 0$, so we can expect that if C is a good fit for D we are near a maximum of the likelihood function. Taking another derivative and averaging, we get:

$$F_{ij=} \langle \mathcal{L}_{,ij} \rangle = \frac{1}{2} Tr \left[A_i A_j \right] \tag{28}$$

Assume we have a single parameter (e.g. Ω). We would like to find a (linear) transformation that will give the most information regarding this parameter. We know that discarding information cannot improve the situation if our parameter space includes the true model, so we assume that we have beforehand decided to transform our n data points into m data points, effected by a transformation $\tilde{d} = \Psi d$ as above. For the case $m = 1$ the optimal choice is a

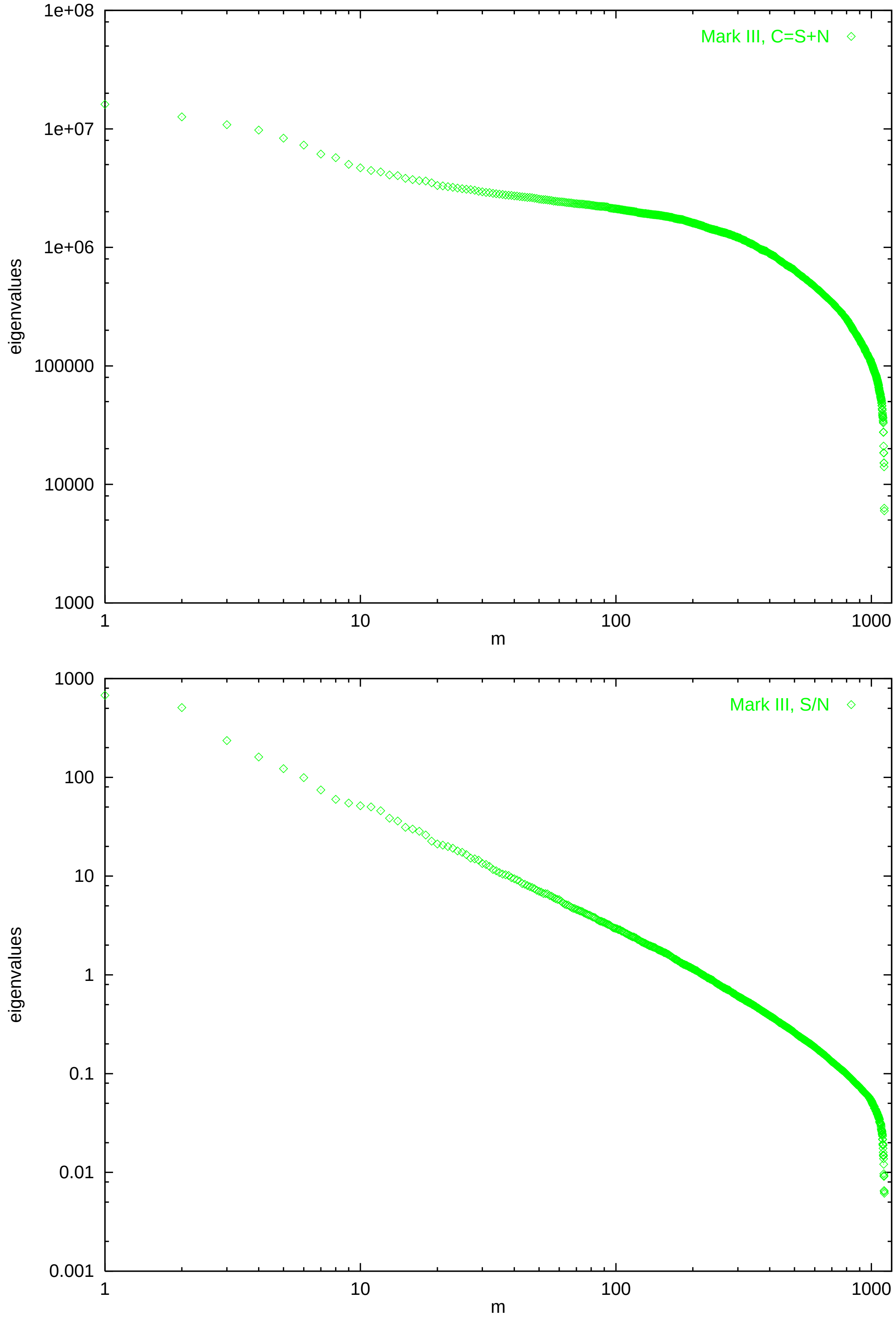


Figure 1: Eigenvalues of the correlation matrices for the data (top) and the whitened data (bottom) for Mark III with ΛCDM power spectrum. For the plot on the right the matrix is: $\tilde{C} = N^{-1/2} S N^{-1/2}$

vector Ψ such that

$$C_{,i}\Psi = \lambda C\Psi \tag{29}$$

This, then, is the generalized eigenvalue equation we must solve to find the best modes. Since the matrices are symmetric we can write $C = RR^\dagger$ and we then have:

$$(R^{-1}C_{,i}R^{\dagger -1})(R^\dagger\Psi) = \lambda(R^\dagger\Psi) \tag{30}$$

Where we can use standard diagonalization techniques since the matrix on the left is symmetric. This generalizes to general m, with each row of Ψ satisfying the equation. If we choose the vectors $R^\dagger\Psi$ normalized to unity then we will have $\left\langle d\tilde{d}\right\rangle = \Psi RR^\dagger\Psi^\dagger = I$ (they are orthogonal as eigenvectors of a symmetric matrix) so these are statistically independent variables. Regarding the error bars, choosing the above normalization gives $\Psi^\dagger C_{,i}\Psi = \Lambda$, with Λ the diagonal matrix of the corresponding eigenvalues, and thus:

$$\tilde{F}_{ii} = Tr\left[\Psi^{\dagger -1}\Lambda^2\Psi^\dagger\right] = \sum_{k=1}^{m}\lambda_k^2 \tag{31}$$

When λ_k are the generalized eigenvalues. This is a formal justification for the choice of the modes corresponding to the *larger* eigenvalues when truncating. We see that in the cast that θ_i is the normalization of the power spectrum, then $C_{,i} = S$, and the above equation is equivalent to $S\Psi = \frac{\lambda}{1-\lambda}\Psi$, or

$$N^{-1/2}SN^{-1/2}(N^{1/2}\Psi) = \frac{\lambda}{1-\lambda}(N^{-1/2}\Psi) \tag{32}$$

which is the 'whitening' transformation discussed later.

2.2 Examination of the Eigenmodes, Whitening

Before representing the measured data in the new basis, it is fruitful to analyze the basis elements themselves. The prior correlation matrix depends on three factors: the prior model and survey geometry for the signal part, and on the survey error estimates for the noise part. Examining the eigenmodes can give information on these factors. This also gives some information on the way this transform reorganizes the data.

A very illustrative example is the eigenmodes of the Mark III survey geometry and error estimates, with the best-fit ΛCDM model from the Zaroubi et. al. analysis. It is easy to ascribe various statistics to the eigenmodes: given some statistic of the survey galaxies $r(g)$ (e.g. radial distance) the average of this statistic for the eigenvector v is:

$$\langle r \rangle_v = \sum |v(g)|^2 r(g) \tag{33}$$

Where the sum is over the survey galaxies (note that the eigenvalues are normalized to unity, so that is also the sum of the weights). We get this way an 'average distance' associated with each mode. In the same way it is possible to define a 'radius variance' for the mode, as $\left\langle (\langle r \rangle_v - r)^2 \right\rangle_v$. If this variance is small with respect to the average distance, we can say that most of the information of this mode comes from galaxies with a certain radial distance. The correlation of eigenvalue with distance is important for two reasons. First of all, the error is proportional to the distance, so more distant modes also are more noisy than nearby ones. Also, as described before, [3] (analyzing the SFI sample) show that dividing the survey around the median distance and analyzing each part separately gives significantly different results for the density parameter. Since we consider truncating the resulting data vector, we need to know if this will introduce (or remove) systematic effects.

Figure 2 clearly shows the distance dependency of the modes[3]. The natural interpretation of this result is that for the correlation matrix: $C = S + N$ except for the first few eigenmodes, the 'N' part dominates the diagonalization. This situation is unfortunate since we cannot simply use the nearby, small-eigenvalue modes alone for further parameter estimation as the error in determining these is large.

A different way of testing the transformation is via the χ^2 goodness-of-fit criterion. On average, the value of $\chi^2 = d^\dagger C^{-1} d$ should be about the sample size n for a Gaussian distribution. Earlier this was used as an indication for the goodness-of-fit of the resulting power spectrum to the measured data. In our case, however, assuming that we know the true correlation matrix for the data, the variables $\widetilde{d}$ are *uncorrelated*. This means that we expect $\chi_i^2 = \frac{\widetilde{d_i^2}}{\lambda_i}$ to be about unity for each mode separately. By testing this premise in various ways we have a finer condition for the goodness-of-fit, in that we both demand a good fit for each mode separately, and also test whether the eigenmodes of the prior correlation matrix actually appear uncorrelated, with variances determined by the eigenvalues. As suggested by Y. Hoffman in [5], we examine the average value $\frac{1}{m}\sum_{i=1}^{m} \chi_i^2$ as a function of m, expecting a value near unity. The normalized standard deviation of a χ^2 distribution with m degrees of freedom is approximately $\sqrt{\frac{2}{m}}$.

Plotting this (figure 3) we see the systematic effect described earlier ('two-halves' problem) - the more distant part of the data (larger eigenvalues here) favours a smaller normalization of the power-spectrum than the nearby part. We can take two points of view with regard to this figure - either as indicating a problem with the physical model (not having a good fit to the power spectrum), or as a indicating problem with the statistical model itself, i.e. with the Gaussian approximation to the signal and the errors.

We will now consider another transformation which will give us a better-fitting correlation matrix in this sense. Returning to the early discussion of a change of parameters, we see that

[3]Note: the eigenmodes are always numbered in order of *decreasing* eigenvalues

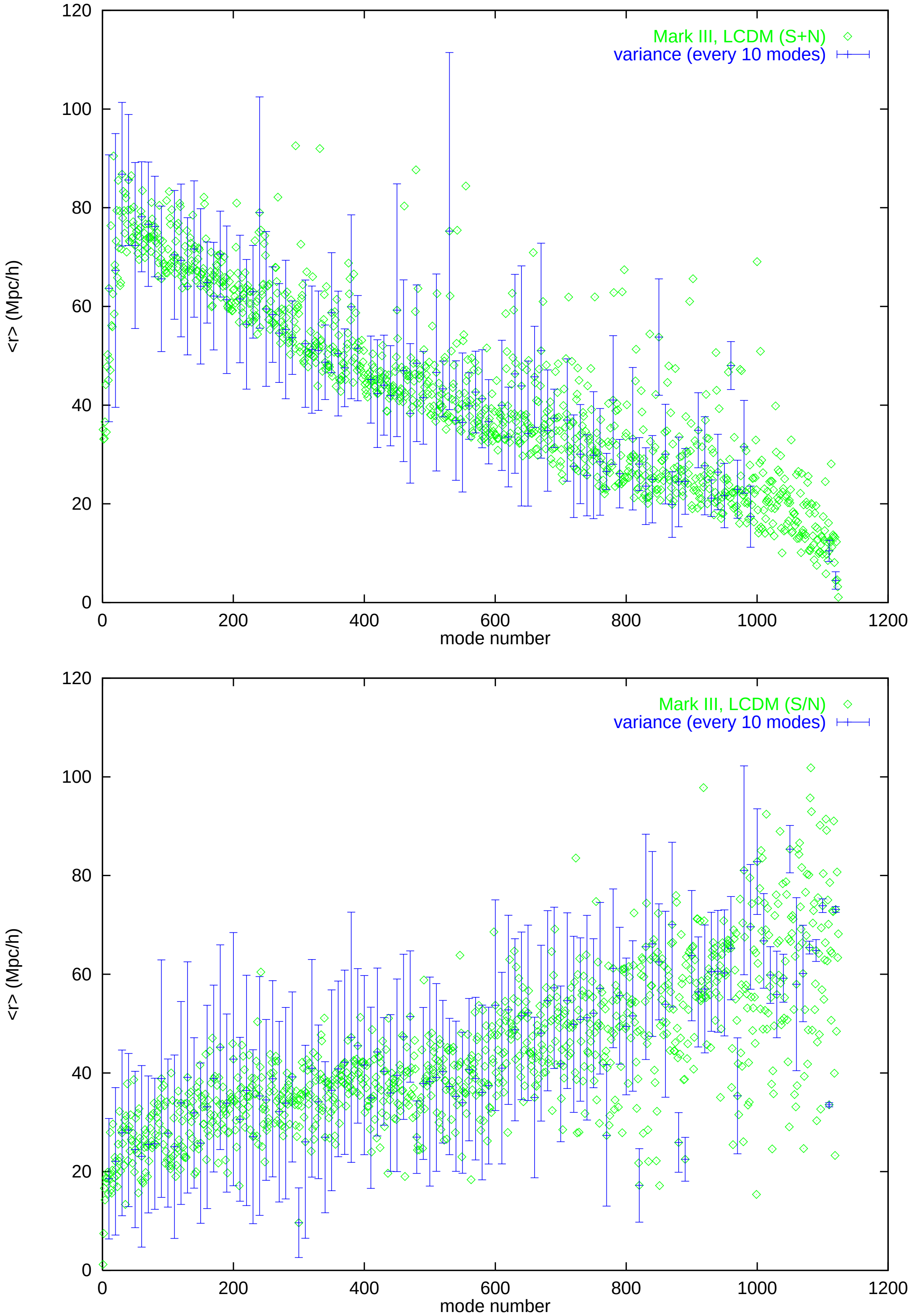
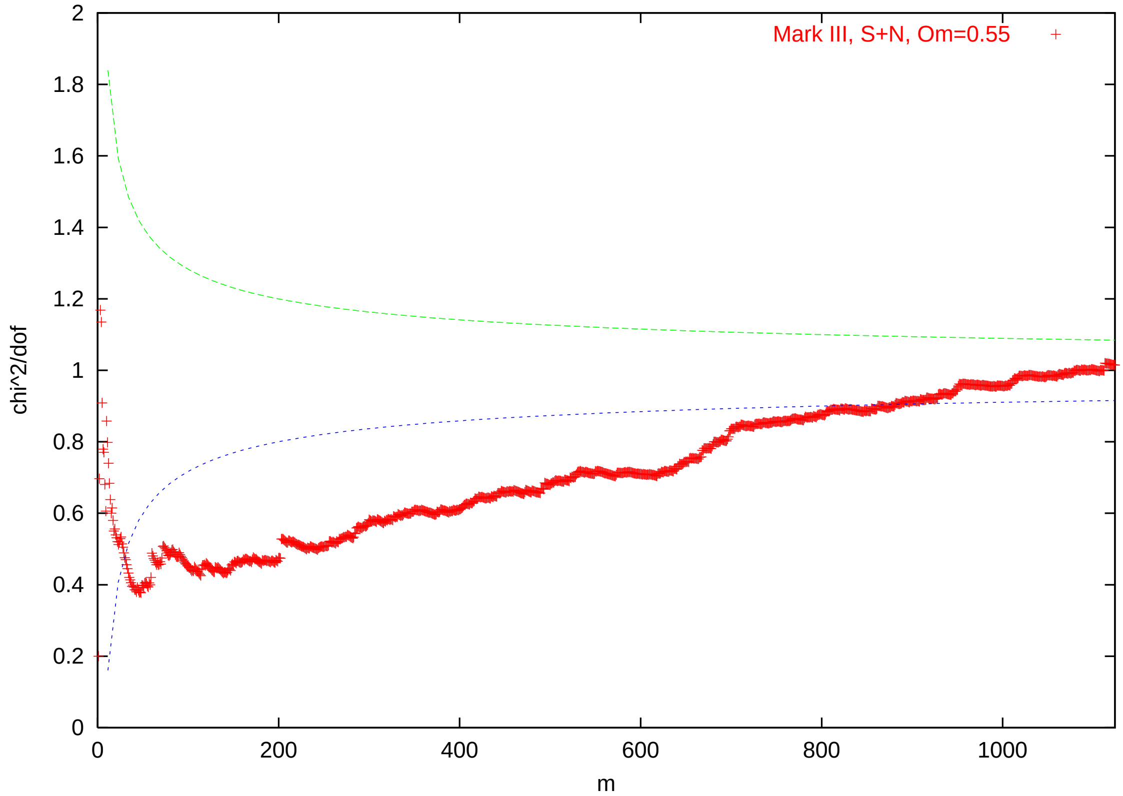


Figure 2: Average distance statistics for the eigenmodes of the correlation matrix for the Mark III catalogue with the ΛCDM power spectrum.



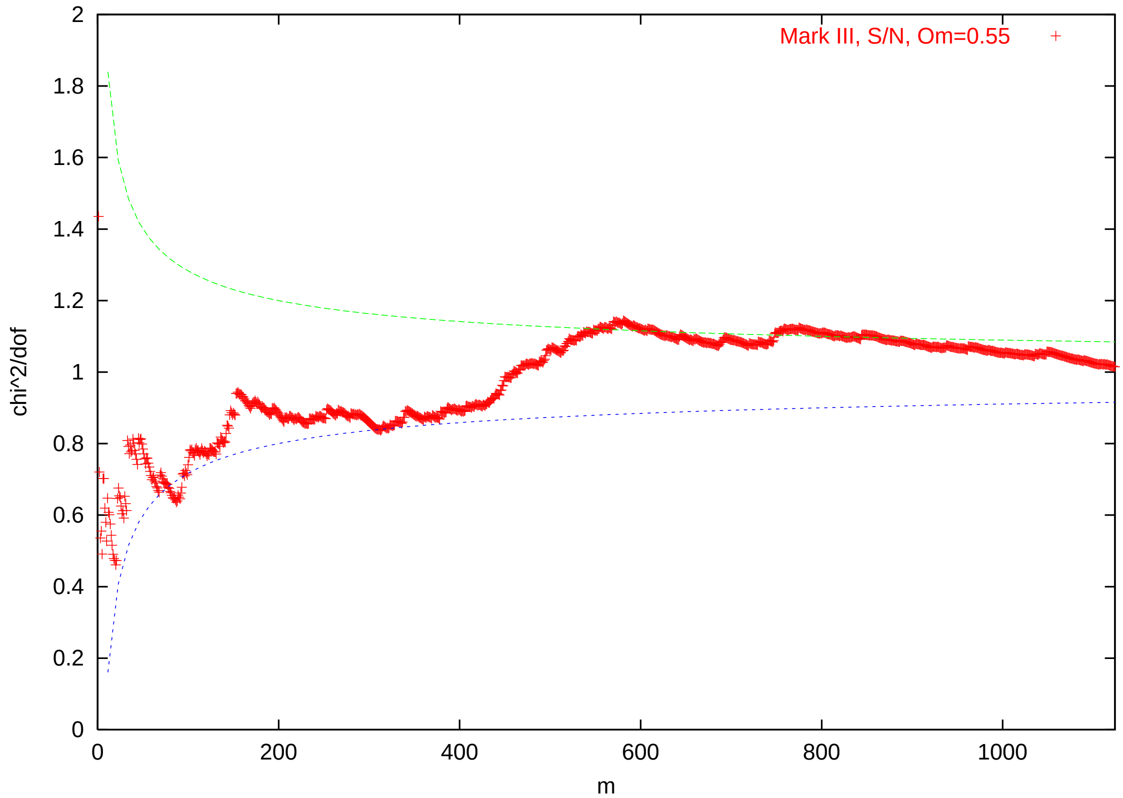


Figure 3: Cumulative $\chi^2/d.o.f.$ plot for the ΛCDM power spectrum (2σ error bars). PCA prior matrix is $C = S + N$ (top) and $\tilde{C} = N^{-1/2} S N^{-1/2}$ (bottom)

we can first make an initial change of variables, $d \to Rd$, to get a new correlation matrix $RCR^\dagger$, and then proceed with the PCA analysis for the new variables and correlation matrix, in the hope that for the new set of variables the analysis will yield more accurate results. A natural choice , "Whitening", suggested by [13], is to define $\hat{d} = N^{-\frac{1}{2}}d$, where N is the noise matrix. This amounts to normalizing the data in terms of the expected noise, in the case of a diagonal (uncorrelated) noise matrix. Representing the correlation matrix as $C = S + N$ we get a new correlation matrix, $\hat{C} = N^{-\frac{1}{2}}SN^{-\frac{1}{2}} + I$, where I is the identity matrix. In the diagonal representation we have a better notion of the signal/noise value of a mode - the eigenvalue (since the 'noise' matrix is I, in this case). Moreover, the larger eigenvalues correspond to larger signal/noise ratios. This does not directly mean that using only some of these modes will lead to better results. If the error model and the velocity distribution model are correct, then since the estimation is asymptotically optimal, using only some of the data can never improve the results (but it will worsen the estimated error bars). Also, if we are keeping all the information (using an invertible transformation matrix R) then the likelihood function will not change.

In our case, however, both the velocity power spectrum model and the error model are suspect for various reasons. Considering the error model first, we note that since the correlation matrix is quadratic in the data, using an incorrect error model can lead to systematic errors in the results. Using modes with a higher signal-to-noise ratio is one of the ways to overcome this difficulty. The eigenvalues for the 'whitened' correlation matrix are plotted in figure 1. Looking at the distance correlation (figure **??**), we see that the correlation is weaker, but it is in the expected direction. We also see that the signal-to-noise modes probe more scales than the normal PCs. The cumulative χ^2 plot (figure 3) also indicates a better fit. Since we are considering the same data and the same model, however, we must justify this claim. The earlier discussion is a real indication that our model is not entirely accurate, but since we saw that most of the information in that analysis came from the noisiest part of the data, the results are suspect.

For a better indication of the agreement for model and data for different parts of the data, it makes sense to group the modes into bins and plot the χ^2 statistic for each bin. Comparing the results for the original correlation matrix and the signal-to-noise matrix (figure 4) the distance dependence problem is clear. The signal-to-noise plot seems to indicate that the problem is not with the most noisy data (since it has little effect on the analysis except when used all by itself), but rather with the data around the median eigenmodes.

3 Broken Power Spectrum

3.1 A Break in the power spectrum

As the last section demonstrated, there are several problems with fitting a ΛCDM power spectrum to the real universe data as is. One of the reasons for this problem may be the fact

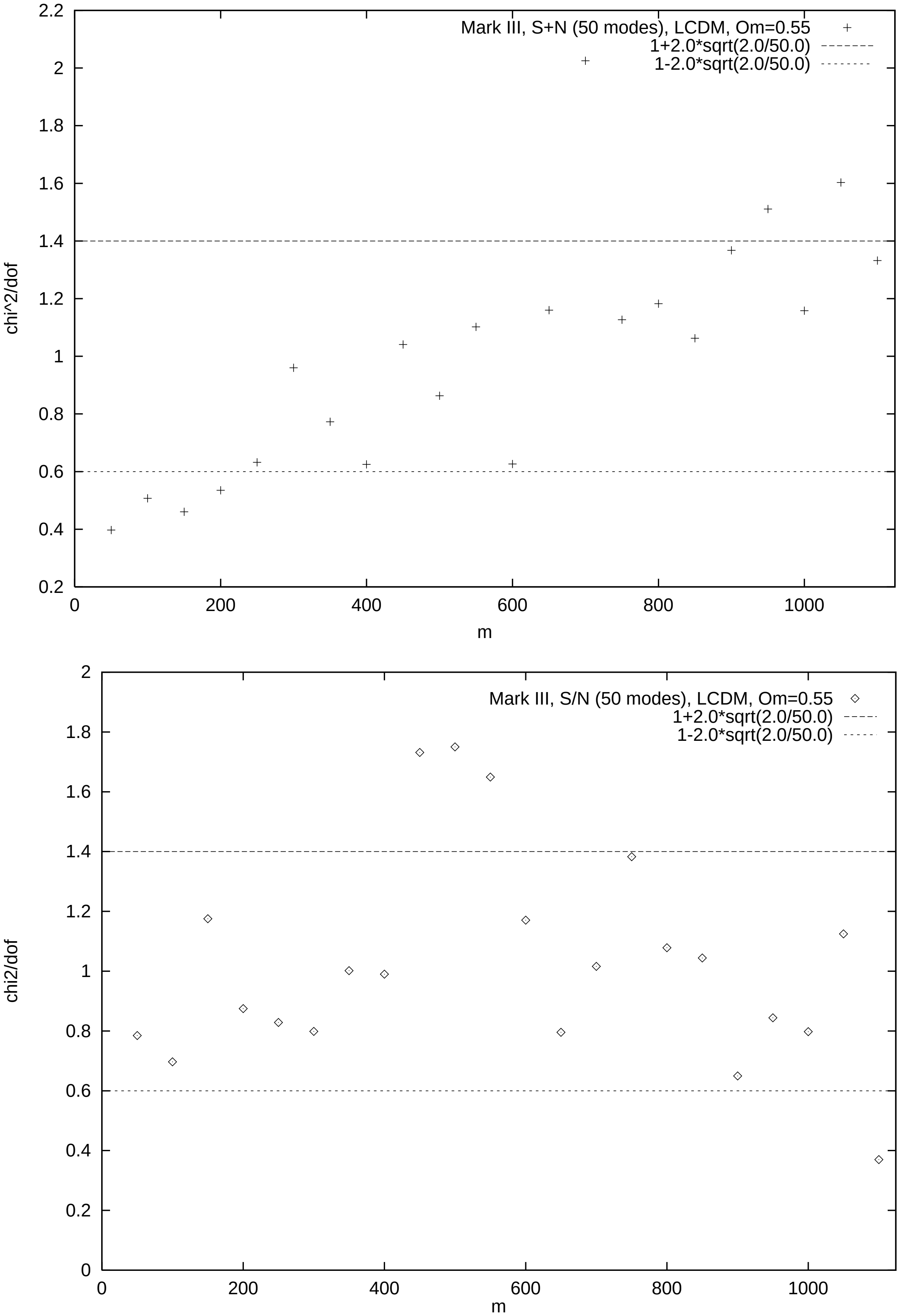


Figure 4: Differential χ^2 plot for Mark III, with the ΛCDM model. Top - C. Bottom - $\hat{C}$

that the nearby data dominates the fit (because it has less noise), thus making the fit strongly dependent on the small-scale structure probed by the data. This is unfortunate, since these are precisely the scales for which our theoretical model, based on the linear approximation, is weakest. For example, the model described earlier (section 1.1) does not account for the fact that as the perturbations grow in contrast, the dense regions also collapse while the voids expand. Thus perturbations on an original comoving scale k_L today correspond to perturbation in a comoving scale k_{NL}. In addition, the shape of the power spectrum changes in a nontrivial way.

In our case, fitting the linear CDM-family models to catalogues reflecting nonlinear effects could lead to systematic errors, since changing the density parameter Ω for these models has more effect on the power spectrum on the smaller scales.

Furthermore, there is some evidence to suggest that there is a 'dip' in the observed power spectrum around the scale of $k \approx 0.1hMpc^{-1}$, a feature missing from the CDM power spectrum. This casts some doubts on the prospects of recovering the cosmological parameters along the lines of the analysis above. Due to the various uncertainties in the theoretical power spectra, we also continue the research along a different line, dropping the goal of estimating the parameters for the simpler goal of estimating the power spectrum. By representing the power spectrum as a linear combination of step functions, the height at each step being a free parameter, it is possible to directly fit the power spectrum to the data without recourse to a specific model. This approach has the additional benefit of being computationally efficient: since the correlation matrix is linear in the power spectrum (see section 1.3), one can calculate the correlation integrals once, and then the correlation matrix is simply a linear combination of the individual steps correlation matrices. This also simplifies the calculation of the derivatives $C_{,i}$.

It would be better, however, to try and combine this suggestion into the existing framework of the model-driven power spectrum, since this might allow to estimate the cosmological parameters, and also generate better-fitting power spectra. This leads naturally to the concept of a 'broken' power spectrum, the name deriving from the resulting shape. Since for the larger scales we consider our theoretical models to provide a good fit to the data, we try a mixed scheme, where the power spectrum at small wavenumbers (large scales) up to a certain point k_b is a CDM variant, and beyond that point it is 'broken' into bands. To approximate the shape of common power spectra, the power spectrum in the smallest-scale band was taken to be a power-law, with both the slope and the normalization left as free parameters. In this way, we can fit the small-scale data while reducing the effect it has on the fits to the cosmological parameters, which we try to extract from the large-scale power spectrum fit. We will present in this section the results of a 2-band scheme: for $k \leq 0.2h\mathrm{Mpc}^{-1}$, the power spectrum of the previously used ΛCDM model, and for $k \geq 0.2h\mathrm{Mpc}^{-1}$ we use $P(k) = Bk^{-s}$.

3.2 Testing the method

For an initial impression regarding the claim that nonlinear effects cause the improvement seen when using a broken power spectrum, we can compare this power with an approximation for a nonlinear power spectrum for the same model. For this we first note that the various power spectra considered were actually approximations for $k^{-2}P_v(k)$ and not for $f(\Omega)P_\delta(k)$. In the linear approximation these two are equal, but if we wish to discuss nonlinear effects the situation is different. For the density power spectrum, Peacock & Dodds in [9] derive an interpolation formula which gives the nonlinear density power spectrum which results from a given linear density power spectrum. Their approximation gives more power on smaller scales, and appears to be compatible with the ad-hoc power law used here (figure 6), especially in the case of SFI. In the absence of a nonlinear approximation to the *velocity* power spectrum, however, this similarity should be considered a hint for further exploration. Freudling et.al ([3]) report that for a certain simulation the linear velocity power spectrum was a better approximation to the nonlinear velocity power spectrum than the power spectrum obtained by using the velocity-density relation and the PD approximation, despite the fact that the latter was a good approximation for the density power spectrum. Turning to newer mock catalgoues, based on the unconstrained GIF simulation, for which the initial power spectrum is ΛCDM, it has become possible to test this again.

The power spectrum for the simulation is difficult to calculate (the geometry of the simulation is complex), but it is possible to compare the results of a direct ΛCDM fit to the a broken-ΛCDM fit in terms of the recovered density parameter (which is known). By carefully constructing the mock catalogues, it is possible to analyze how the presence of nonlinear scales in the catalogue affects the result. 100 mock catalgoues, made to resemble Mark III in the selection effects and bias corrections were drawn from the simulation (ΛCDM, $\Omega = 0.3$): 10 samples of galaxies from the simulation were used. As in the original Mark III catalogue, galaxies near the center of a rich cluster are counted as a single data point (with the average peculiar velocity), which reduces both the errors and the nonlinear effects. Other cluster galaxies are considered 'field' galaxies. This should not be confused with the grouping process done later. In the real-universe catalogue the center of a cluster included, on average, the spiral galaxies up to a radius of $3.5h^{-1}$Mpc and the elliptical galaxies up to a radius of $1.5h^{-1}$Mpc. By varying the radius used to identify the centers in the mock catalogue (10%,20%, ... ,100% of the originals), one has 10 versions of the catalogue for each sample, each version showing a different level of nonlinear effects.

For the catalogues with the largest radii, fitting with the ΛCDM model resulted in an average fit of $\Omega \approx 0.35 \pm 0.04$, which is close to the true result. Fitting a broken ΛCDM power spectrum gave a slightly better value of $\Omega \approx 0.33 \pm 0.05$ which is not significantly different. Also, the likelihood improvements were quite marginal ($\Delta \ln \mathcal{L} \approx 0.5$ in most cases). For the catalogues with the most severe nonlinear effects, the results were dramatic: fitting the original ΛCDM power spectrum gave an average fit of $\Omega \approx 0.68 \pm 0.05$, while fitting a broken power

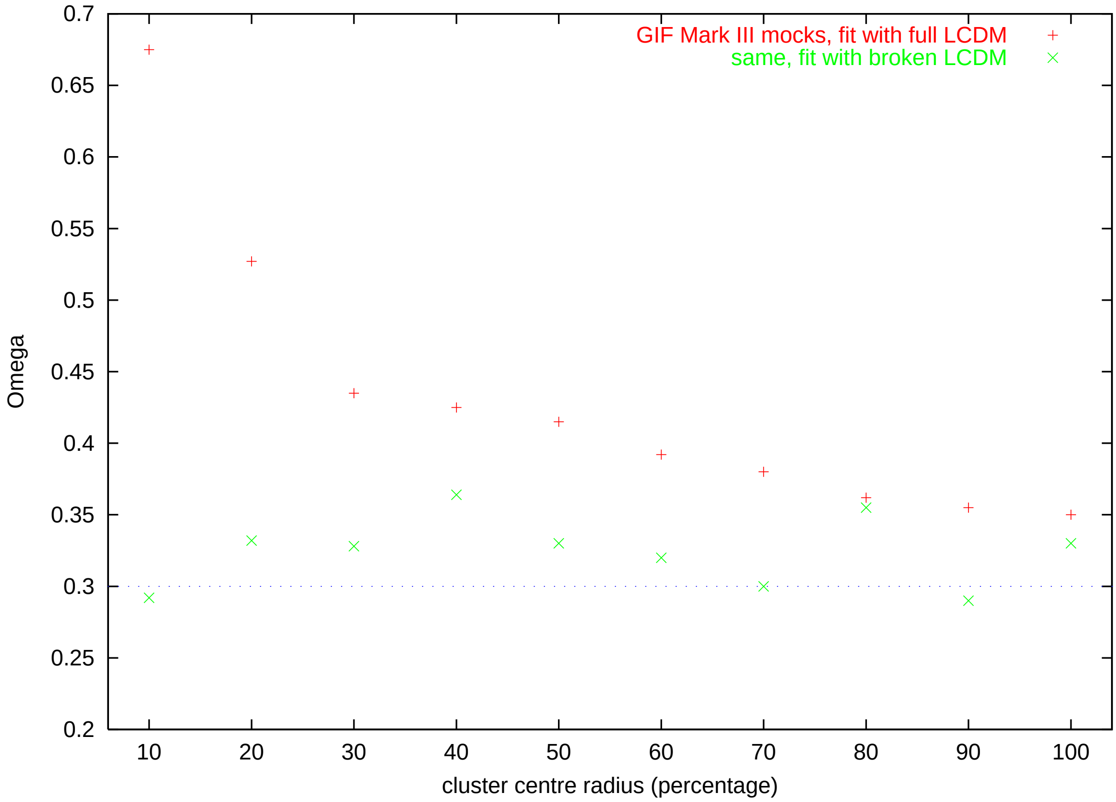


Figure 5: Recovering the simulation density parameter (actual value $\Omega = 0.3$) using the ΛCDM and broken ΛCDM power spectra, from 10 sets of mock catalogues showing different levels of nonlinear effects

spectrum gave $\Omega \approx 0.29 \pm 0.06$. The likelihood gain was also significant ($\Delta \ln \mathcal{L} \approx 120$). Figure 5 shows the dependence of the recovered results on the construction of the data.

3.3 The Real Universe

In figure 6 we show the best-fit power spectra for both catalogues, compared with the original analysis. We see that the large-scale data indeed prefers significantly lower values for the density parameter, $\Omega \approx 0.32 \pm 0.05$ For Mark III and $\Omega \approx 0.37 \pm 0.08$ for SFI. For Mark III the likelihood improvement is dramatic ($\Delta \ln \mathcal{L} \sim 22$), while for SFI it is significant, but modest ($\Delta \ln \mathcal{L} \sim 2.8$). In terms of the likelihood function, the probability density around the Mark III data for the ΛCDM power spectrum is 9.5 *orders of magnitude* smaller than for the broken model. The broken model fit to the SFI catalogue is found to be 'only' 16 times more likely than the ΛCDM fit. The value of $\sigma_8 \Omega^{0.56}$ is ~ 0.51 for Mark III, and ~ 0.66 for SFI which are closer to the cluster normalization value of 0.57 quoted earlier. We also see that the data prefer much more power at lower scales than the CDM model has. This leads to a better fit in the PCA χ^2 tests (figure 7), where we see a weaker dependence on the distance.

The 'broken' power spectrum also has a new parameter - the break location k_b. It is possible to subject this parameter to the likelihood analysis, but this is doubtful since we are not sure of the model. We would like to make sure, however, that the analysis is robust: in the limit

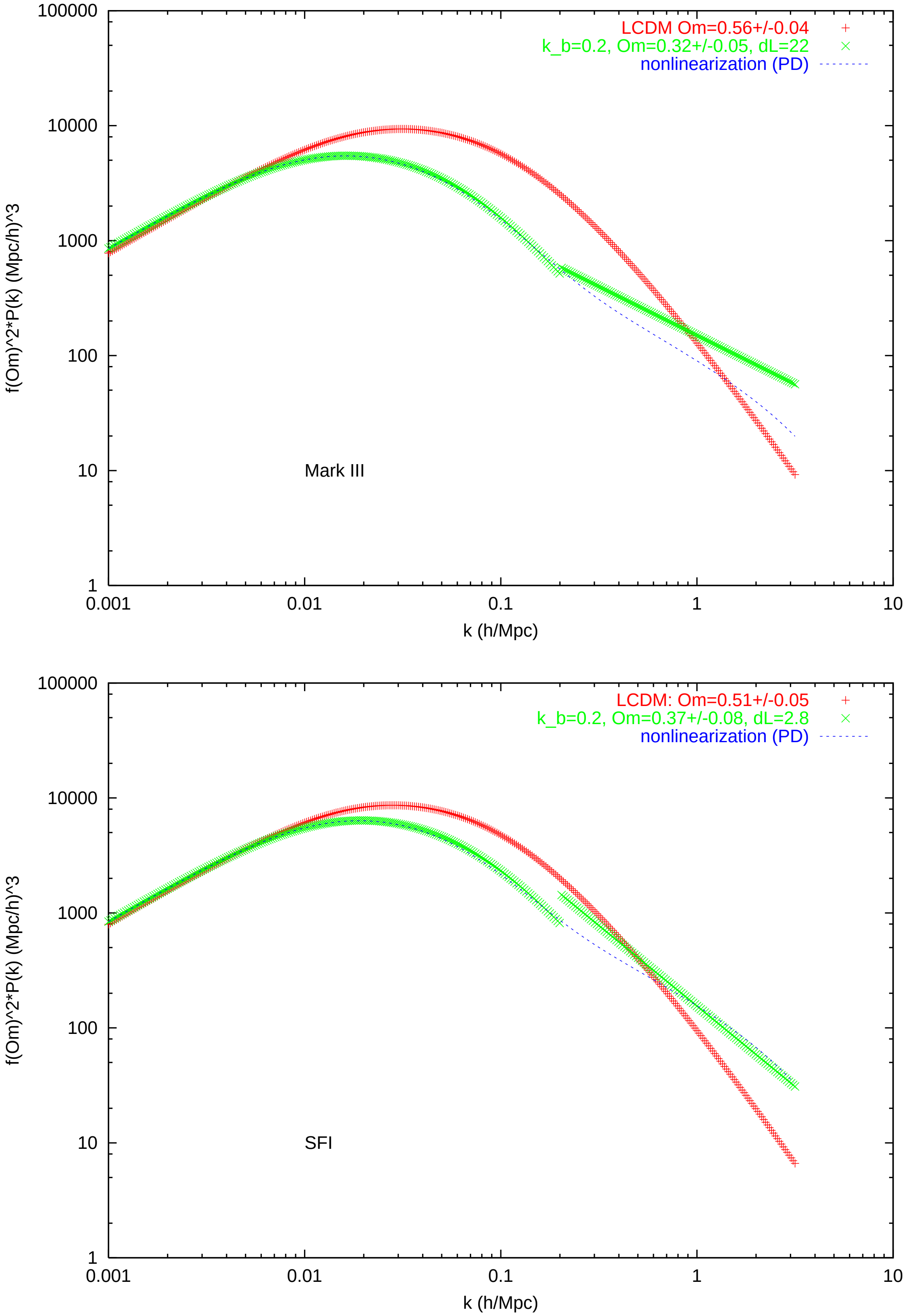


Figure 6: ΛCDM power spectra with and without a break, and the PD nonlinear power spectra calculated for the linear ΛCDM recovered to the left of the break for the Mark III (top) and SFI (bottom) catalogues.

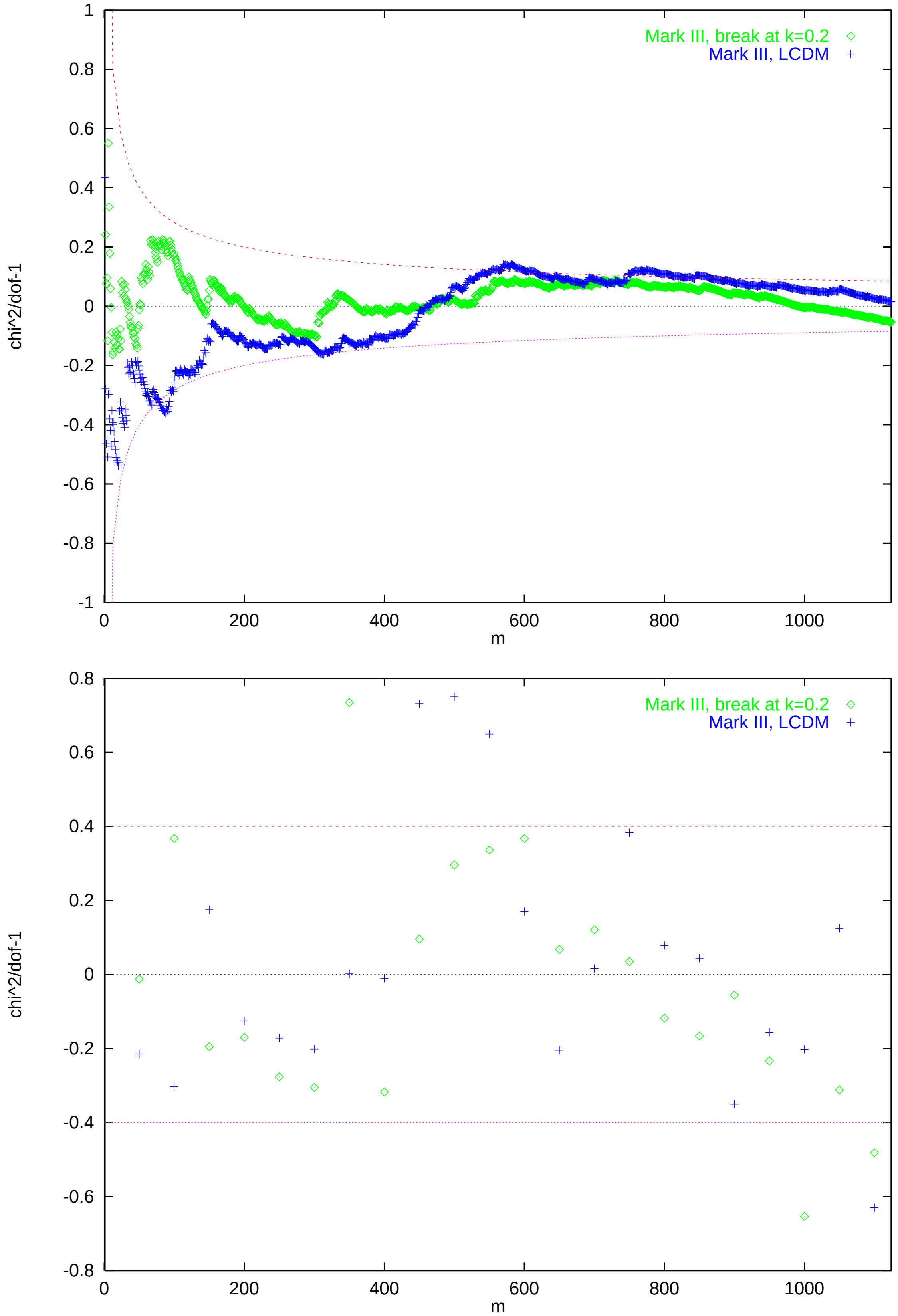


Figure 7: χ^2 tests comparing the original ΛCDM fit to the 'broken' model (2σ error bars)

of large values of k_b we should recover the results of the original analysis. Figure 8 shows the situation for the real universe catalgoues. If we break the power spectrum at very large scales, we cannot hope for significant results since the data used cannot constrain the power spectrum, well in these regions. We also see that the recovered density parameter does not significantly depend on the break location for most of the range considered. When we move the break to smaller scales we see, however, that we only return to the non-broken case for a very high break location near the 'infinity' of the numerical integration. We do not have a satisfactory explanation for this behaviour.

3.4 The 4-band model and the 'dip'

In addition to the 'broken' model discussed above, a model with 4 bands was tested, with a surprising result. The best-fit 2-band model power spectra appear continuous (see figure 6). On the other hand, when two additional bands (taken as step functions) were added, one for $0.02 \leq k \leq 0.07$, and one for $0.07 \leq k \leq 0.2$ (in units of $h\mathrm{Mpc}^{-1}$), the best-fit power spectra appeared different (figure 9). The significant likelihood gain, especially for Mark III, is probably due to the power law on the small scales, and not to the additional freedom in the medium scales. The most striking feature is that in both cases, the analysis prefer much less power in the $0.07 \leq k \leq 0.2$ band than in the CDM model. A hint for similar feature has been observed in the galaxy distribution power spectrum of the APM survey ([1]). This feature may indicate the presence of an HDM (massive neutrinos) or a baryonic component in the total mass, which were ignored in the previous analysis, but may also be due to cosmic variance.

4 Conclusion

Given the 'correct' model, likelihood analysis is the (statistically) best method to recover the cosmological parameters. This work presents tools to recover the parameters using incomplete theoretical knowledge (error model, nonlinear phenomena). PCA gives an understanding of the behaviour of the model and the relation between the data and the resulting best-fit model. We used this to analyze the 'two-halves' behaviour of the previous results, both by analyzing both the geometric properties ($< r >$) and statistics (χ^2 tests) of the fit on a mode-by-mode basis. We found that while the correlation matrix C did not describe the data well, the 'whitened' correlation matrix $N^{-1/2}CN^{-1/2}$ provided a better description of the 'whitened' data, $N^{-1/2}d$. We also saw that in the case of the 'whitened' data, the fit is governed by data on the smaller scales, which is more prone to nonlinear effects.

PCA also allows one to use the parts of the data which best determine a given parameter. This compression is useful if the model is incomplete, mandatory if the data is excessively large. The basic formalism was discussed, but the important question of the point of truncation was not addressed. Applying likelihood analysis to larger samples will required some solution to this problem.

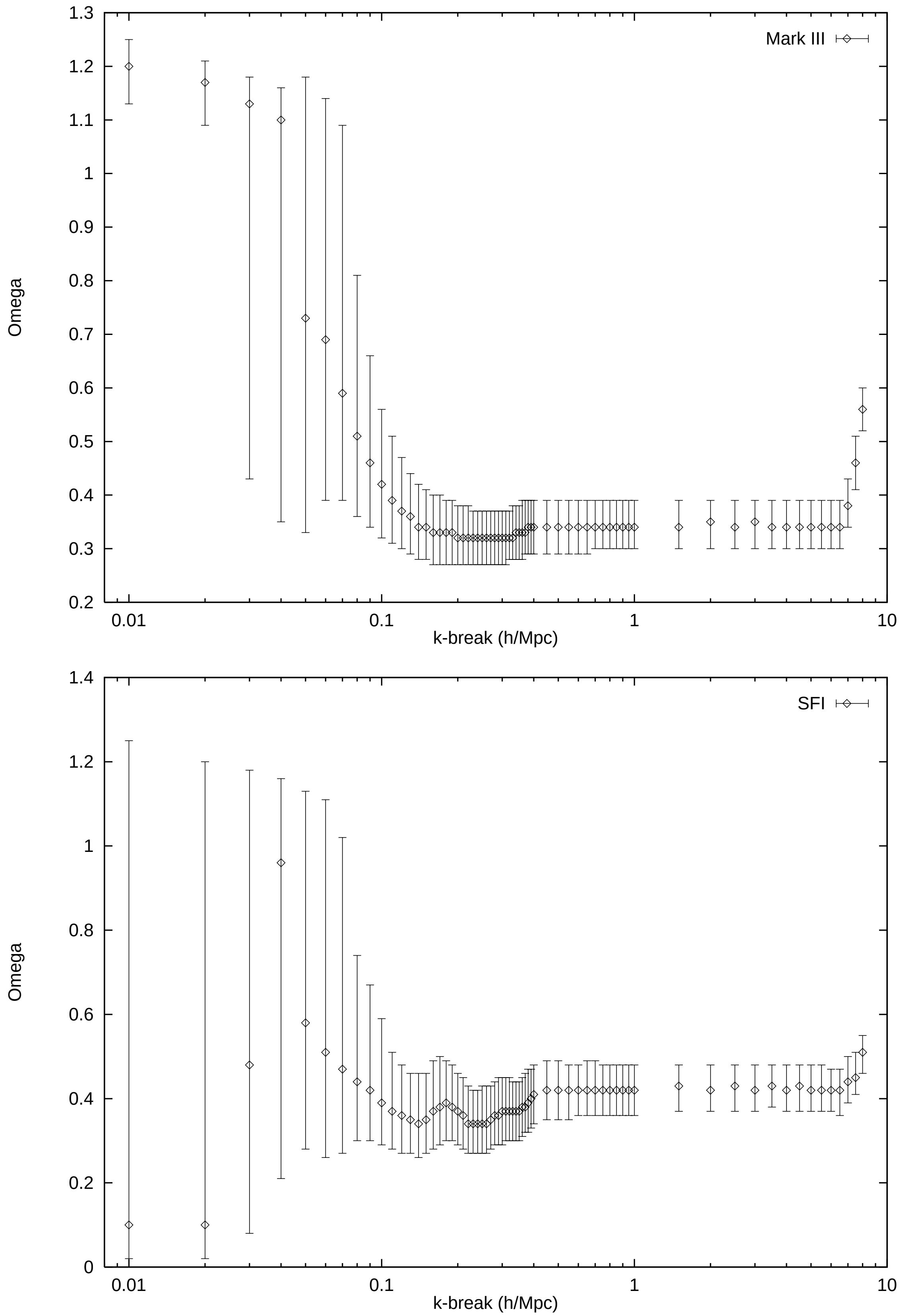


Figure 8: Best-fit value of Ω as a function of break location, for Mark III (top) and SFI (bottom)

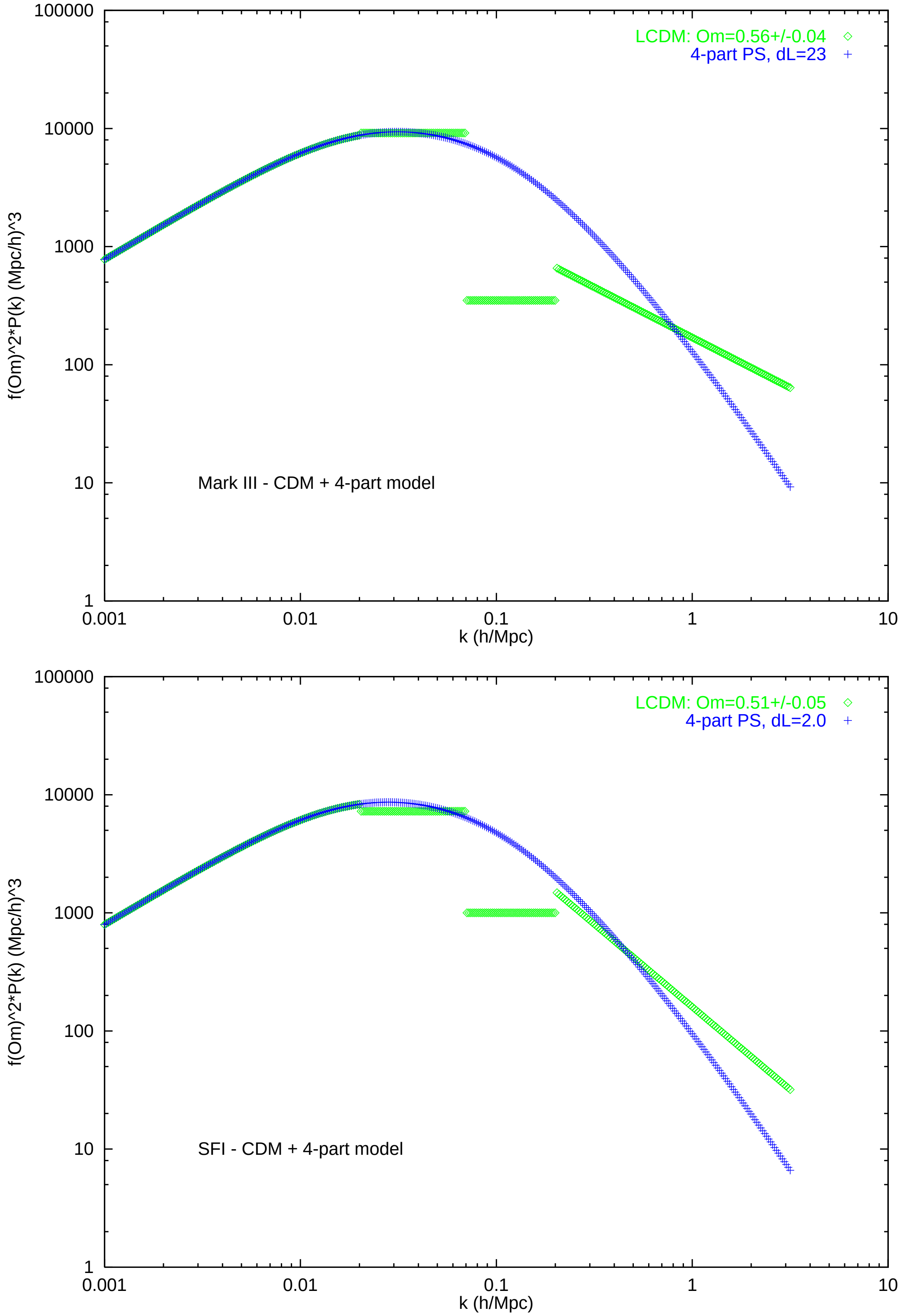


Figure 9: ΛCDM power spectra and a 4-band power spectrum for Mark III (top) and SFI (bottom) catalogues

Breaking the power spectrum up into bands allows an efficient way to determine the power spectrum, but usually precludes the recovery of the cosmological parameters. We have shown that using a physically motivated model (ΛCDM in our case) for one of the bands allows the determination of the parameters. This combined approach fits the strong areas of the model to the data while reducing the effects of the regions where this model is weak. After validating the method using mock catalogues, we use it to fit the Mark III and SFI catalgoues and find $\Omega \approx 0.32/0.37$ and $\sigma_8\Omega^{0.56} \approx 0.51/0.66$ respectively, which are in greater agreement with other cosmological studies than the previous results of Zaroubi et. al. and Freudling et. al. The 'broken' model also appears to be a better fit according to the PCA χ^2 tests. Testing with mock catalogues indicate that the improvement with respect to the Mark III catalogue is due to better handling of nonlinear effects.

Using the 'broken' models we have also observed a 'dip' feature in the fitted power spectrum. This may be due to cosmic variance within the framework of the ΛCDM model, or an essential feature of the power spectrum, perhaps indicating a significant amount of HDM/baryonic mass.

Acknowledgements

I thank Prof. Avishai Dekel for much teaching and guidance. I would also like to thank Prof. Yehuda Hoffman and Prof. Ofer Lahav for stimulating discussions and suggestions, Dr. Idit Zehavi for the use of the source code from her programs and Amiram Eldar for creating the mock catalgoues and assisting with their use.

References

[1] Baugh, C.M. & Gaztañaga, E., 1998, in *Proceedings: Evolution of Large Scale Structure* (astro-ph/9810184)

[2] Eisenstein, D. & Hu, W., 1999, ApJ 511, p. 5

[3] Freudling W., Zehavi, I., da Costa, L.N., Dekel, A., Eldar A., Giovanelli, R., Haynes, M.P., Salzer, J.J., Wegner, G., Zaroubi, S., 1999, ApJ 523, p. 1

[4] Haynes, M.P., Giovanelli, R., Chamaraux, P., da Costa, L.N., Freudling W., Salzer, J.J., Wegner, G., 1999, AJ 117, p. 2039

[5] Hoffman, 1999, preprint (astro-ph/9990158)

[6] Kauffmann, G., Colberg, J., Diaferio, A. & White, S.D.M., 1999, MNRAS 303, p. 188

[7] Kolatt, T., Dekel, A., Ganon, G. & Willick, J.A., 1996, ApJ 458, p. 419

[8] Lahav, O., Lilje, P.B., Primack, J.R., Rees, M.J., 1991, MNRAS 251, p. 128

[9] Peacock, J.A. & Dodds, S.J., 1996, MNRAS 280, p. L19

[10] Peebles, P.J.E., 1980, The Large-Scale Structure of the Universe, Princeton Univ. Press, Princeton, NJ

[11] Sugiyama, N., 1995, ApJ (Supp.) 100, p. 281

[12] Tegmark, M., Taylor, A.N. & Heavens, A.F., 1997, ApJ 480, p. 22

[13] Vogeley, M.S. & Szalay, A.S., 1996, ApJ 465, p. 34

[14] White, S.D.M., Efstathiou, G., & Frenk, C.S., 1993, MNRAS 262, p. 1023

[15] Willick, J.A., Courteau, S., Faber, S.M., Burstein, D. & Dekel, A., 1995, ApJ 446, p. 12

[16] Willick, J.A., Courteau, S., Faber, S.M., Burstein, D., Dekel, A. & Kolatt, T., 1996, ApJ 457, p. 460

[17] Willick, J.A., Courteau, S., Faber, S.M., Burstein, D., Dekel, A. & Strauss, M., 1997, ApJ (Supp.) 109, p. 333

[18] Zaroubi S., Zehavi, I., Dekel, A., Hoffman, Y. & Kolatt, T., 1997, ApJ 486, p. 21