

# Bayesian power-spectrum inference for Large Scale Structure data

Jens Jasche<sup>1</sup>, Francisco S. Kitaura<sup>2,1</sup>, Benjamin D. Wandelt<sup>3</sup>, Torsten A. Enßlin<sup>1</sup>

<sup>1</sup> *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild Straße 1, D-85748 Garching, Germany*

<sup>2</sup> *SISSA, Scuola Internazionale Superiore di Studi Avanzati, via Beirut 2-4, I-34014 Trieste, Italy*

<sup>3</sup> *Department of Physics University of Illinois at Urbana-Champaign, 1110 West Green Street Urbana, IL 61801-3080, USA*

Submitted to MNRAS 22-May-2009

## ABSTRACT

We describe an exact, flexible, and computationally efficient algorithm for a joint estimation of the large-scale structure and its power-spectrum, building on a Gibbs sampling framework and present its implementation ARES (Algorithm for REconstruction and Sampling). ARES is designed to reconstruct the 3D power-spectrum together with the underlying dark matter density field in a Bayesian framework, under the reasonable assumption that the long wavelength Fourier components are Gaussian distributed. As a result ARES does not only provide a single estimate but samples from the joint posterior of the power-spectrum and density field conditional on a set of observations. This enables us to calculate any desired statistical summary, in particular we are able to provide joint uncertainty estimates. We apply our method to mock catalogs, with highly structured observational masks and selection functions, in order to demonstrate its ability to reconstruct the power-spectrum from real data sets, while fully accounting for any mask induced mode coupling.

**Key words:** large scale – reconstruction – Bayesian inference – cosmology – observations – methods – numerical

## 1 INTRODUCTION

Throughout cosmic history a wealth of information on the origin and evolution of our Universe has been imprinted to the large scale structure via the gravitational amplification of primordial density perturbations. Harvesting this information from probes of the large scale structure, such as large galaxy surveys, therefore is an important scientific task to further our knowledge and to establish a conclusive cosmological picture. In recent years great advances have been made, both in retrieving huge amounts of data and increasing sensitivity in galaxy redshift surveys. Especially the recent galaxy surveys, the 2dF Galaxy Redshift Survey (Colless et al. 2001) and the Sloan Digital Sky Survey (Abazajian et al. 2008) provide sufficient redshifts to probe the 3D galaxy distribution on large scales. In particular, the two point statistics of the matter distribution contains valuable information to test standard inflation and cosmological models, which describe the origin and evolution of all observed structure in the Universe. Measuring the power-spectrum from galaxy observations therefore has always attracted great interest. Precise determination of the overall shape of the power-spectrum can for instance place important constraints on neutrino masses, help to identify the primordial power-spectrum, and break degeneracies for cosmological parameter estimation from CMB data (e.g. Hu et al. 1998; Spergel et al. 2003; Hannestad 2003; Efstathiou et al. 2002; Percival et al. 2002; Spergel et al. 2003; Verde et al. 2003). In addition, several characteristic length scales have been imprinted to the matter distribution throughout cosmic history, which can serve as new standard rulers to measure the Uni-

verse. A prominent example of these length scales is the sound horizon, which yields oscillatory features in the power-spectrum, the so called baryon acoustic oscillations (BAO) (e.g. Silk 1968; Peebles & Yu 1970; Sunyaev & Zeldovich 1970). Since the physics governing these oscillatory features is well understood, precise measurements of the BAO will allow us to establish a new, precise standard ruler to measure the Universe through the distance redshift relation (Blake & Glazebrook 2003; Seo & Eisenstein 2003). Precision analysis of large scale structure data therefore is a crucial step in developing a conclusive cosmological theory.

Unfortunately, contact between theory and observations cannot be made directly, since observational data is subject to a variety of systematic effects and statistical uncertainties. Such systematics and uncertainties arise either from the observational strategy or are due to intrinsic clustering behavior of the galaxy sample itself (Sánchez & Cole 2008). Some of the most prominent uncertainties and systematics are:

- survey geometry and selection effects
- close pair incompleteness due to fiber collisions
- galaxy biases
- redshift space distortions

The details of galaxy clustering, and how galaxies trace the underlying density field are in general very complicated. The bias between galaxies and mass density is most likely non-linear and stochastic, so that the estimated galaxy spectrum is expected to differ from that of the mass (Dekel & Lahav 1999). Even in the limit where a linear bias could be employed, it still differs for dif-

ferent classes of galaxies (see e.g. Cole et al. 2005). In addition, the apparent density field, obtained from redshift surveys, will generally be distorted along the line-of-sight due to the existence of peculiar velocities.

However, the main cause for the systematic uncertainties in large scale power-spectrum estimations is the treatment of the survey geometry (Tegmark 1995; Ballinger et al. 1995). Due to the survey geometry the raw power-spectrum yields an expectation value for the power-spectrum, which is the true cosmic power-spectrum convolved with the survey mask (Cole et al. 2005). This convolution causes an overall distortion of the power-spectrum shape, and drastically decreases the visibility of the baryonic features.

The problems, mentioned above, have been discussed extensively in literature, and many different approaches to power-spectrum analysis have been proposed. Some of the main techniques to recover the power-spectrum from galaxy surveys are Fourier transform based, such as the optimal weighting scheme, which assigns a weight to the galaxy fluctuation field, in order to reduce the error in the estimated power (see e.g. Feldman et al. 1994; Tegmark 1995; Hamilton 1997a; Yamamoto 2003; Percival et al. 2004). Alternative methods rely on Karhunen-Loève decompositions (Tegmark et al. 1997; Tegmark & et al. 2004; Pope et al. 2004) or decompositions into spherical harmonics, which is especially suited to address the redshift space distortions problematic (Fisher et al. 1994; Heavens & Taylor 1995; Tadros et al. 1999; Percival et al. 2004; Percival 2005). In addition, to these deconvolution methods there exists a variety of likelihood methods to estimate the real space power-spectrum (Ballinger et al. 1995; Hamilton 1997a,b; Tadros et al. 1999; Percival 2005). In order to not just provide the maximum likelihood estimate but also conditional errors, Percival (2005) proposed a Markov Chain Monte Carlo method to map out the likelihood surface.

Nevertheless, as the precision of large scale structure experiments has improved, the requirement on the control and characterization of systematic effects, as discussed above, also steadily increases. It is of critical importance to propagate properly the uncertainties caused by these effects through to the matter power-spectrum and cosmological parameters estimates, in order to not underestimate the final uncertainties and thereby draw incorrect conclusions on the cosmological model.

We therefore felt inspired to develop a new Bayesian approach to extract information on the two point statistics from a given large scale structure dataset. We prefer Bayesian methods to conventional likelihood methods, as they yield more general and profound statements about measurements. In example, conventional likelihood methods can only answer questions of the inner form like :” Given the true value  $s$  of a signal, what is the probability distribution of the measured values  $d$ ?” A Bayesian method, on the other hand, answers questions of the type :” Given the observations  $d$ , what is the probability distribution of the true underlying signal  $s$ ?” For this reason, Bayesian statistics answers the underlying question to every measurement problem, of how to estimate the true value of the signal from observations, while conventional likelihood methods do not (Michel & Kirchhoff 1999). Since the result of any Bayesian method is a complete probability distribution they permit fully global analyses, taking into account all systematic effects and statistical uncertainties. In particular, here, we aim at evaluating the power-spectrum posterior distribution  $\mathcal{P}(\{P(k_i)\}|d_i)$ , with  $P(k_i)$  being the power-spectrum coefficients of the  $k_i$ th mode and  $d_i = d(\vec{x}_i)$  is an observation at position  $\vec{x}_i$  in three dimensional configuration space. This probability distribution would then contain all information on the two point statistics

supported by the data. In order to explore this posterior distribution we employ a Gibbs sampling method, previously applied to CMB data analysis (see e.g. Wandelt 2004; Eriksen et al. 2004; Jewell et al. 2004).

Since direct sampling from  $\mathcal{P}(\{P(k_i)\}|d_i)$  is impossible or at least difficult, they propose instead to draw samples from the full joint posterior distribution  $\mathcal{P}(\{P(k_i), \{s_i\}|d_i)$  of the power-spectrum coefficients  $P(k_i)$  and the 3D matter density contrast amplitudes  $s_i$  conditional on a given set of data points  $\{d_i\}$ . This is achieved by iteratively drawing density samples from a Wiener-posterior distribution and power-spectrum samples via an efficient Gibbs sampling scheme (see figure 1 for an illustration). Here, artificial mode coupling, as introduced by survey geometry and selection function, is resolved by solving the Wiener-filtering equation, which naturally regularizes inversions of the observational response operator in unobserved regions. In this fashion, we obtain a set of Monte Carlo samples from the joint posterior, which allows us to compute any desired property of the joint posterior density, with the accuracy only limited by the sample size. In particular, we obtain the power spectrum posterior  $\mathcal{P}(\{P(k_i)\}|d_i)$  by simply marginalizing the joint posterior  $\mathcal{P}(\{P(k_i), \{s_i\}|d_i)$  over the auxiliary density amplitudes  $s_i$ , which is trivially achieved by ignoring the  $s_i$  samples.

The Gibbs sampler also offers unique capabilities for propagating systematic uncertainties end-to-end. Any effect, for which we can define a sampling algorithm, either jointly with or conditionally on other quantities, can be propagated seamlessly through to the final posterior.

It is worth noting, that our method differs from traditional methods of analyzing galaxy surveys in a fundamental aspect. Traditional methods consider the analysis task as a set of steps, each of which arrives at intermediate outputs which are then fed as inputs to the next step in the pipeline. Our approach is a truly global analysis, in the sense that the statistics of all science products are computed jointly, respecting and exploiting the full statistical dependence structure between various components.

In this paper we present ARES (Algorithm for REconstruction and Sampling), a computer algorithm to perform a full Bayesian data analysis of 3D redshift surveys. In section 3 we give an introduction to the general idea of the large scale structure Gibbs sampling approach, followed by section 4 and 5, where we describe and derive in detail the necessary ingredients to sample the 3D density distribution and the power-spectrum respectively. The choice of the prior and the relevance for the cosmic variance are discussed in section 6. Details concerning the numerical implementation are discussed in section 7. We then test ARES thoroughly in section 8, particularly focussing on the treatment of survey masks and selection functions. In section 9 we demonstrate the running median filter, and use it as an example to demonstrate how uncertainties can be propagated to all inferences based on the set of Gibbs samples. Finally we conclude in section 10, by discussing the results of the method and giving an outlook for future extensions and application of our method.

## 2 NOTATION

In this section, we describe the basic notation used throughout this work. Let the quantity  $\rho_i = \rho(\vec{x}_i)$  be the field amplitude of the three dimensional field  $\rho(\vec{x})$  at position  $\vec{x}_i$ . Then the index  $i$  has to be understood as a multi index, which labels the three components of the position vector:

$$\vec{x}_i = [x_i^1, x_i^2, x_i^3], \quad (1)$$

where  $x_i^j$  is the  $j$ th component of the  $i$ th position vector. Alternatively one can understand the index  $i$  as a set of three indices  $\{r, s, t\}$  so that for an equidistant grid along the three axes the position vector can be expressed as:

$$\vec{x}_i = \vec{x}_{r,s,t} = [\Delta x r, \Delta y s, \Delta z t], \quad (2)$$

with  $\Delta x$ ,  $\Delta y$  and  $\Delta z$  being the grid spacing along the three axes. With this definition we yield:

$$\rho_i \equiv \rho_{r,s,t}. \quad (3)$$

Also note that any summation running over the multi index  $i$  is defined as the three sums over the three indices  $r, s$  and  $t$ :

$$\sum_i \equiv \sum_r \sum_s \sum_t. \quad (4)$$

Further, we will frequently use the notation  $\{\rho_i\}$ , which denotes the set of field amplitudes at different positions  $\vec{x}_i$ . In particular:

$$\{\rho_i\} \equiv \{\rho_0, \rho_1, \rho_2, \dots, \rho_{N-1}\}, \quad (5)$$

where  $N$  is the total number of position vectors.

### 3 THE LARGE SCALE STRUCTURE GIBBS SAMPLER

As already described in the introduction, we seek to sample from the joint posterior distribution  $\mathcal{P}(\{P(k_i)\}, \{s_i\}|\{d_i\})$  of the power-spectrum coefficients  $P(k_i)$  and the 3D matter density contrast amplitudes  $s_i$  given a set of observations  $\{d_i\}$ .

In principle, this joint posterior distribution could be mapped out over a grid in the multi-dimensional space of the signal amplitudes  $s_i$  and power-spectrum coefficients  $P(k_i)$ . But since the number of grid points required for such an analysis scales exponentially with the number of free parameters, this approach cannot be realized efficiently. For this reason, we propose a Gibbs sampling approach to this problem.

The theory of Gibbs sampling (Gelfand & Smith 1990; Tanner 1996; O'Hagan 2000) states, that if it is possible to sample from the conditional densities  $\mathcal{P}(\{s_i\}|\{P(k_i)\}, \{d_i\})$  and  $\mathcal{P}(\{P(k_i)\}|\{s_i\}, \{d_i\})$ , then iterating the following two sampling equations will, after an initial burn-in period, lead to samples from the joint posterior  $\mathcal{P}(\{s_i\}, \{P(k_i)\}|\{d_i\})$ :

$$\{s_i\}^{(j+1)} \sim \mathcal{P}(\{s_i\}|\{P(k_i)\}^{(j)}, \{d_i\}), \quad (6)$$

$$\{P(k_i)\}^{(j+1)} \sim \mathcal{P}(\{P(k_i)\}|\{s_i\}^{(j+1)}, \{d_i\}), \quad (7)$$

where the symbol  $\sim$  denotes a random draw from the probability density on its right.

Once a set of samples from  $\mathcal{P}(\{s_i\}, \{P(k_i)\}|\{d_i\})$  has been obtained, the properties of this probability density can be summarized in terms of any preferred statistic, such as its multivariate mean, mode or variance.

As our approach probes the joint distribution, we are able to quantify joint uncertainties of the signal amplitudes and the power-spectrum conditional just on the data. For this reason, the Gibbs sampling approach should not be considered as yet another maximum likelihood technique, although it certainly is able to produce such an estimate.

In the following we are going to describe the necessary methods and procedures required for iterating the processes 6 and 7 of signal and power-spectrum sampling.

## 4 SAMPLING THE SIGNAL MAPS

Assuming a Gaussian signal posterior distribution  $\mathcal{P}(\{s_i\}|\{P(k_i)\}, \{d_i\})$ , the task of drawing a random signal sample can be split into two steps.

First, we estimate the maximum a posteriori values for the signal amplitudes  $s_i$ , which in the Gaussian case coincide with the mean values. Then a fluctuation term, being consistent with the correct covariance, is added to the mean signal. The sum of the mean and the fluctuation term will then yield a sample from the conditional posterior.

The most challenging procedure in this signal sampling step is to calculate the a posteriori values for the signal amplitudes  $s_i$ . Assuming a Gaussian posterior will directly lead to a Wiener filtering procedure, described below. However, this method requires to invert huge matrices which consists of the sum of the inverse signal  $S$  and inverse noise  $N$  covariance matrices. The matrix inversion is a numerically very demanding step, and at the same time presents the bottleneck for our method, as it defines the computational speed with which a signal sample can be produced. The efficient implementation of these matrix inversion step, as described by Kitaura & Enßlin (2008), allows for the production of many thousands of samples in computational feasible times. In the following sections we will describe the details of the signal sampling procedure.

### 4.1 The Wiener filter

As already described above, the main task for the signal sampling step is to derive the maximum a posteriori values for the signal amplitudes  $s_i$ . According to Bayes' theorem the conditional signal posterior can be written as the product of a signal prior and a likelihood normalized by the so called evidence. Further, here we will use the signal covariance matrix  $S$  rather than the power-spectrum  $\{P(k_i)\}$ . It is well known, that the power-spectrum is just the Fourier transform of the signal covariance in configuration space. Since the Fourier transform is a basis transformation with a unitary transformation matrix, the signal covariance matrix  $S$  and the power-spectrum  $\{P(k_i)\}$  can be used interchangeably for a normalized Fourier transform (see section 5.1 and Appendix B for more details).

We can therefore write the signal posterior as:

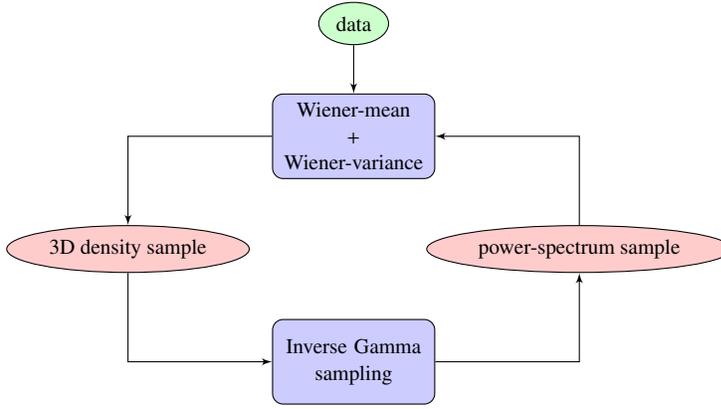
$$\begin{aligned} \mathcal{P}(\{s_i\}|S, \{d_i\}) &= \frac{\mathcal{P}(S)}{\mathcal{P}(\{d_i\}, S)} \mathcal{P}(\{s_i\}|S) \mathcal{P}(\{d_i\}|\{s_i\}, S) \\ &= \frac{1}{\mathcal{P}(\{d_i\}|S)} \mathcal{P}(\{s_i\}|S) \mathcal{P}(\{d_i\}|\{s_i\}), \end{aligned} \quad (8)$$

where we assume that the data amplitudes  $d_i$  are conditionally independent of the signal covariance matrix  $S$ , once the signal amplitudes  $s_i$  are given. Following Bardeen et al. (1986), we describe the signal prior for the large scale matter distribution as a multivariate Gaussian, with zero mean and the signal covariance  $S$ . We can then write:

$$\mathcal{P}(\{s_i\}|S) = \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{1}{2} \sum_i \sum_j s_i S_{ij}^{-1} s_j}. \quad (9)$$

The Fourier transform of the signal covariance matrix  $S$  has an especially appealing form in Fourier space. It is well known, that in an homogeneous and isotropic universe the Fourier transform of the signal covariance is a diagonal matrix, with the diagonal elements being the power-spectrum. Hence, we can express the Fourier representation of the signal covariance as:

$$\hat{S}_{kl} = \delta_{kl}^K P_k \quad (10)$$



**Figure 1.** Flow-chart depicting the two step iterative Gibbs sampling procedure.

where the  $\hat{\cdot}$ -symbol denotes a Fourier transform,  $\delta_{ij}^k$  is the Kronecker delta and  $P_k = P(k_k)$  is the power-spectrum coefficient at the Fourier mode  $\vec{k}_k$  in three dimensional pixel space (see e.g. Padmanabhan 1993; Lahav & Suto 2004).

The choice of the Gaussian prior can be justified by inflationary theories, which predict the matter field amplitudes to be Gaussian distributed in the linear regimes at scales  $k \lesssim 0.15$  h/Mpc (Peacock & Dodds 1994; Percival et al. 2001).

At nonlinear scales the Gaussian prior does not represent the full statistical behavior of the matter field anymore. During nonlinear gravitational structure formation the statistics of the initial density field has evolved from a Gaussian distribution towards a log normal distribution as commonly assumed in literature (Coles & Jones 1991; Colombi 1994; Kayo et al. 2001).

However, note that in this case the Gaussian prior still describes the two point statistics of the underlying density field even in the nonlinear regime. The Gaussian prior should therefore be regarded as our a priori knowledge of the matter distribution, which is only formulated up to two point statistics. Next, we discuss the likelihood  $\mathcal{P}(\{d_i\}|\{s_i\})$  given in equation (8).

As we seek to recover the maximum a posteriori signal  $s_i$  from the set of observations  $d_i$  we must assume a model which relates these both quantities. The most straight forward data model is linear, and can be written as:

$$d_i = \sum_k K_{ik} s_k + \epsilon_i, \quad (11)$$

where  $K_{ij}$  is an observation response operator and  $\epsilon_i$  is an additive noise contribution, which will be defined in more detail in the next section. If we assume the noise  $\epsilon_i$  to be Gaussian distributed, with zero mean and covariance  $N$ , we can express the likelihood as:

$$\mathcal{P}(\{d_i\}|\{s_i\}) = \frac{1}{\sqrt{\det(2\pi N)}} e^{-\frac{1}{2}(\sum_i \sum_j [d_i - \sum_k K_{ik} s_k] N_{ij}^{-1} [d_j - \sum_l K_{jl} s_l])}, \quad (12)$$

where we simply inserted the data model given in equation (11) into the Gaussian noise distribution.

With these definitions the signal posterior is a multivariate Gaussian distribution and can be written as:

$$\mathcal{P}(\{s_i\}|\mathcal{S}, \{d_i\}) \propto e^{-\frac{1}{2}(\sum_i \sum_j s_i S_{ij}^{-1} s_j + [\sum_k K_{ik} s_k] N_{ij}^{-1} [d_j - \sum_l K_{jl} s_l])}, \quad (13)$$

where we omitted the normalization factor, which is of no interest in the following.

Note, that omitting the normalization of the likelihood, requires that the additive noise term is independent of any signal con-

tribution, as otherwise the noise covariance matrix would carry signal information and could not be neglected in the following. This assumption, however, is in general not true for the Poissonian noise, as described, in the next section.

The maximum of this signal posterior can then easily be found by either completing the square in the exponent of equation (13), or by extremizing  $\mathcal{P}(\{s_i\}|\mathcal{S}, \{d_i\})$  with respect to the signal amplitudes  $s_i$ . The latter approach allows us to directly read of the Wiener filter equation from equation (13), by simply differentiating the exponent with respect to  $s_i$  and setting the result to zero. The result is the famous Wiener filter equation which is given as:

$$\sum_j \left[ S_{ij}^{-1} + \sum_m \sum_l K_{mi} N_{ml}^{-1} K_{lj} \right] m_j = \sum_m \sum_l K_{mi} N_{ml}^{-1} d_l, \quad (14)$$

where we denoted the variable  $m_j$  as a Wiener mean signal amplitude, to clarify that this reconstruction is the mean and not a typical sample of the distribution. The solution of this equation requires to invert the matrix:

$$D_{ij} = S_{ij}^{-1} + \sum_m \sum_l K_{mi} N_{ml}^{-1} K_{lj}, \quad (15)$$

which leads to the solution for the signal amplitudes

$$m_i = \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} d_l. \quad (16)$$

This result demonstrates that estimating the maximum a posteriori values  $m_i$  for the signal amplitudes  $s_i$  involves inversions of the Wiener filter operator  $D$ . Therefore, the signal-sampling operation is by far the most demanding step of our Gibbs sampling scheme, as it requires the solution of a very large linear system.

Formally speaking, in practice, this corresponds to inverting matrices of order  $\sim 10^6 \times 10^6$  or larger, which clearly is not computationally feasible through brute-force methods. For example, matrix inversion algorithms, based on usual linear algebra methods, have a numerically prohibitive  $\mathcal{O}(N_{pix}^3)$  scaling, in order to transform to the eigenspace of the system, which bars sampling from the signal posterior.

This is the situation in which Kitaura & Enßlin (2008) proposed a particular operator based inversion technique to allow for computationally efficient calculation of the Wiener filter equation in three dimensional space.

In this implementation, the system of equations (16) can be solved by means of conjugate gradients (CGs). The computational scaling of this method is thus reduced to the most expensive step for applying the operator on the right-hand side of the equations,

which in our case is the Fast Fourier transform, which scales as  $O(N_{pix} \log(N_{pix}))$ .

## 4.2 The galaxy data model

In order to adapt the Wiener filter procedure for the specific application to galaxy observations, we are going to present the galaxy data model together with the according Poissonian noise covariance matrix.

It is possible to describe the observed galaxy distribution as a realization of an inhomogeneous Poissonian process (Martínez & Saar 2002). We can therefore assume the observed galaxy numbers  $N_i^O = N^O(\vec{x}_i)$  at position  $\vec{x}_i$  in three dimensional configuration space to be drawn from a Poissonian distribution (Martínez & Saar 2002; Kitaura et al. 2009).

$$N_i^O \sim \mathcal{P}(N_i^O | \lambda_i^O) = \frac{\lambda_i^{O N_i^O} e^{-\lambda_i^O}}{N_i^{O!}}, \quad (17)$$

where the arrow denotes a random draw from the probability distribution and  $\lambda_i^O$  is the mean observable galaxy number at position  $\vec{x}_i$ . We can then write the observed galaxy numbers at discrete positions as:

$$N_i^O = \langle N_i^O \rangle + \epsilon_i^O = \lambda_i^O + \epsilon_i^O, \quad (18)$$

where the noise term  $\epsilon_i^O$  denotes the difference between the observed galaxy number and the mean observable galaxy number. The Poissonian noise covariance matrix can then easily be obtained by:

$$N_{ij}^P = \langle \epsilon_i^O \epsilon_j^O \rangle = \langle [N_i^O - \langle N_i^O \rangle][N_j^O - \langle N_j^O \rangle] \rangle = \delta_{ij}^K \langle N_i^O \rangle = \delta_{ij}^K \lambda_i^O, \quad (19)$$

where we simply calculated the Poissonian variance for the observed galaxy number assuming the galaxies to be independent and identically distributed (i.i.d). The mean observable galaxy number can be related to the true mean galaxy number  $\lambda_i$  by applying the observation response operator  $R_{ij}$  as:

$$\lambda_i^O = \sum_j R_{ij} \lambda_j, \quad (20)$$

The true mean galaxy number, on the other hand, can be related to the dark matter over density field, the signal  $s_i$ , by introducing a physical model in the form of a bias operator  $B_{ij}$ , e.g. a scale dependent bias:

$$\lambda_i = \bar{\lambda} \left( 1 + \sum_j B_{ij} s_j \right). \quad (21)$$

By inserting equations (20) and (21) into equation (18) and applying trivial algebraic conversions, we yield the data model:

$$d_i = \frac{N_i^O}{\bar{\lambda}} - \sum_j R_{ij} = \sum_j R_{ij} \sum_k B_{jk} s_k + \frac{\epsilon_i^O}{\bar{\lambda}}, \quad (22)$$

For the case of galaxy redshift surveys the response operator  $R_{ij}$  is the product of the sky mask and the selection function, which are both local in configuration space, and hence the response operator turns to:

$$R_{ij} = \delta_{ij}^K M_i F_i, \quad (23)$$

where  $M_i$  is the value of the sky mask and  $F_i$  is the value of the selection function at position  $i$ . We therefore arrive at the data model already described in equation (11), which reads:

$$d_i = M_i F_i \sum_k B_{ik} s_k + \frac{\epsilon_i^O}{\bar{\lambda}}$$

$$= \sum_k K_{ik} s_k + \epsilon_i, \quad (24)$$

where we introduced the effective observation response operator  $K_{ij} = M_i F_i B_{ij}$  and the noise contribution  $\epsilon_i = \epsilon_i^O / \bar{\lambda}$ . This is the galaxy data model which we derived from the assumption of the Poissonian distribution of galaxies.

The Wiener filter operator requires the definition of the noise covariance matrix  $N$ , which for the Poissonian noise can be expressed as:

$$N_{ij} = \langle \epsilon_i \epsilon_j \rangle = \frac{\langle \epsilon_i^O \epsilon_j^O \rangle}{\bar{\lambda}^2} = \delta_{ij}^K \frac{\lambda_i^O}{\bar{\lambda}^2}, \quad (25)$$

where we used the Poissonian noise covariance matrix given in equation (19).

However, introducing equation (20) and (21) yields the noise covariance matrix:

$$N_{ij} = \delta_{ij}^K \frac{1}{\bar{\lambda}} \left[ \sum_k R_{ik} \left( 1 + \sum_l B_{kl} s_l \right) \right], \quad (26)$$

which immediately reveals, that there is a correlation between the underlying signal amplitudes  $s_i$  and the level of shot noise produced by the discrete distribution of galaxies (see e.g. Seljak 1998).

Nevertheless, as pointed out in the previous section, the Wiener filter relies on the fact, that the additive noise contribution is uncorrelated with the signal. Hence, we have to assume the noise covariance as uncorrelated with the signal, but it may have some structure.

Therefore, we provide two approaches to effectively approximate the noise covariance matrix given in equation (26).

In the first approach we calculate an effective noise covariance matrix by averaging the noise covariance matrix given in equation (26) over the signal. We then obtain:

$$\begin{aligned} \bar{N}_{ij} &= \langle N_{ij} \rangle_s \\ &= \delta_{ij}^K \frac{1}{\bar{\lambda}} \left[ \sum_k R_{ik} \left( 1 + \sum_l B_{kl} \langle s_l \rangle_s \right) \right] \\ &= \delta_{ij}^K \frac{1}{\bar{\lambda}} \left[ \sum_k R_{ik} \right], \end{aligned} \quad (27)$$

where we used the fact, that the ensemble mean of the signal amplitudes for the density contrast vanishes. Note, that this model also arises when pursuing a least squares approach to matter field reconstructions rather than the Bayesian approach as described in this work (for details see Kitaura & Enßlin 2008).

In the other approach we introduce a noise structure function  $n_i^{SF}$  given as:

$$n_i^{SF} = \frac{\lambda_i^O}{\bar{\lambda}^2}. \quad (28)$$

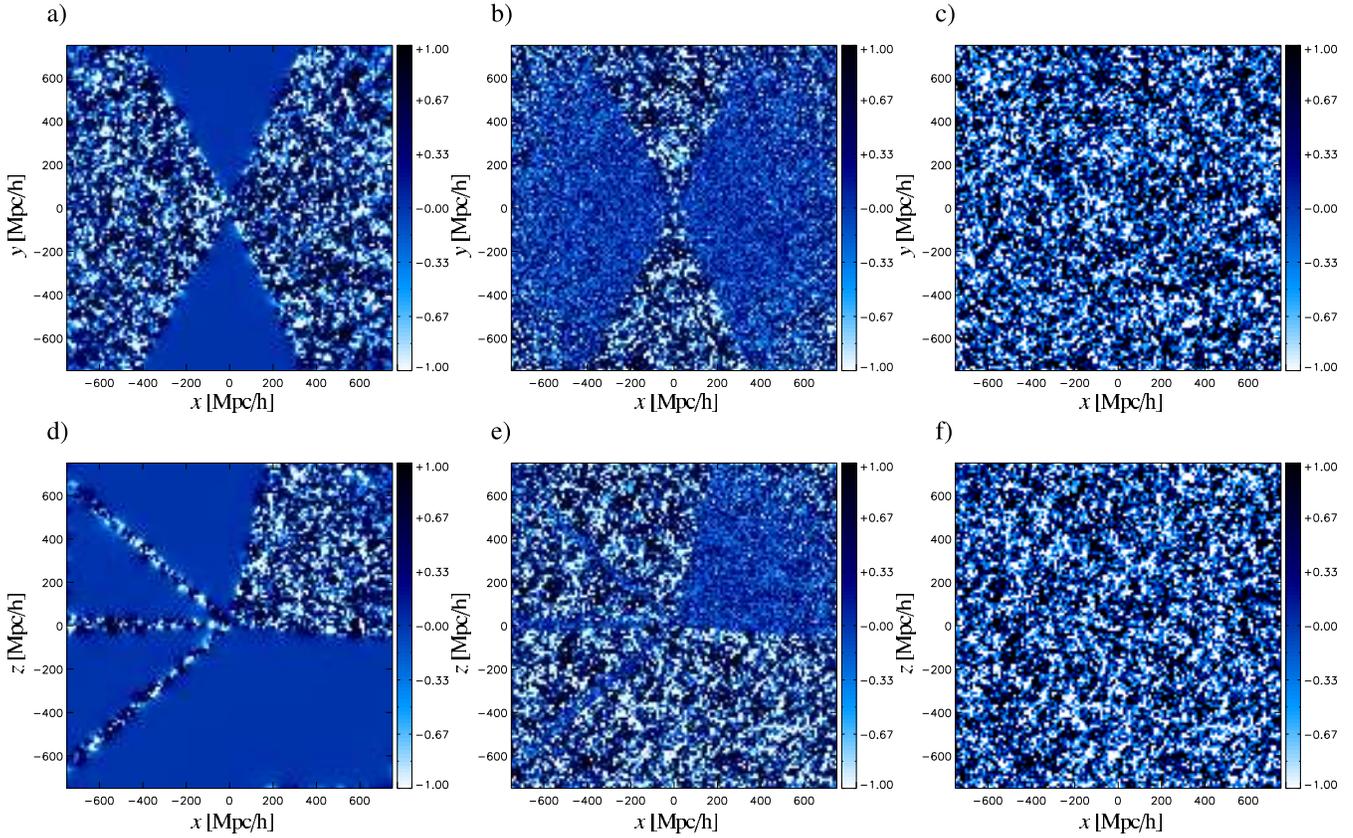
With this definition the noise is approximated as being uncorrelated to the signal, but nonuniform. The noise covariance matrix then reads:

$$N_{ij}^{SF} = \delta_{ij}^K n_i^{SF}. \quad (29)$$

In order to use this noise structure function we have to estimate  $\lambda_i^O$  from the observed galaxy numbers  $N_i^O$ . By applying Bayes' theorem to the Poissonian distribution given in relation (39) we yield:

$$\mathcal{P}(\lambda_i^O | N_i^O) = \mathcal{P}(\lambda_i^O) \frac{\mathcal{P}(N_i^O | \lambda_i^O)}{\mathcal{P}(N_i^O)}. \quad (30)$$

In the absence of any further a priori knowledge on  $\lambda_i^O$  we assume a flat prior and search for the maximum of:



**Figure 2.** Slices through signal sample produced in one Gibbs sampling step. The left panels (a,d) show the Wiener filtered mean signal, panels (b,e) present the fluctuation term, and the right panels show the full, noiseless Gibbs sample. The color map encodes the amplitude of the density contrast.

$$\mathcal{P}(\lambda_i^O | N_i^O) = \frac{\lambda_i^{O N_i^O} e^{-\lambda_i^O}}{\Gamma(N_i^O + 1)}, \quad (31)$$

which is normalized to yield unity when integrated over all  $\lambda_i^O$ .

The noise structure function  $n_i^{SF}$  can then be estimated by searching the most likely value for  $\lambda_i^O$  from equation (31). This yields:

$$n_i^{SF} = \frac{N_i^O}{\lambda^2}. \quad (32)$$

Another estimator for  $n_i^{SF}$  is based on evaluating the mean of the probability distribution given in equation (31). The ensemble mean is calculated as:

$$\langle \lambda_i^O \rangle = \int_0^\infty d\lambda_i^O \lambda_i^O \frac{\lambda_i^{O N_i^O} e^{-\lambda_i^O}}{\Gamma(N_i^O + 1)} = N_i^O + 1. \quad (33)$$

in this case the noise structure function  $n_i^{SF}$  can be written as:

$$n_i^{SF} = \frac{N_i^O + 1}{\lambda^2}. \quad (34)$$

A more thorough discussion on Poissonian noise models and their numerical implications for matter field reconstructions can be found in Kitaura et al. (2009).

### 4.3 Drawing signal samples

In the previous sections, we described the Wiener filter and the galaxy data model, which are required to estimate the mean of

the signal posterior. However, this mean signal is no sample from the signal posterior yet, neither does it represent a physical density field, as it lacks power in the low signal to noise regions. To create a true sample from the signal posterior, one must therefore add a fluctuation term  $y_i$ , which compensates the power lost due to noise filtering. The signal sample can then be written as the sum of the signal mean and a fluctuation term:

$$s_i = m_i + y_i. \quad (35)$$

In our approach we realize the fluctuation term by generating a mock signal  $s_i^*$  and a mock observation  $d_i^*$  consistent with the data model given in equation (24). This kind of mock observation generation is well known in literature and has been applied to various scientific applications, as for instance the generation of constrained initial conditions for Nbody simulations (see e.g. Bertschinger 1987; Hoffman & Ribak 1991; Ganon & Hoffman 1993; Kitaura & Enßlin 2008). The fluctuation term can then simply be calculated as:

$$y_i = s_i^* - \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} d_l^*. \quad (36)$$

The interpretation of this equation is simple. In the high signal to noise regime, the Wiener filter is nearly a pass-through operator, meaning the reconstructed signal is nearly identical to the true underlying signal. Therefore, as the variance is low, the fluctuation term tends towards zero. In the low signal to noise regime, on the other hand, the Wiener filter will block, and no signal can be reconstructed. The fluctuation term will therefore be nearly identical to the underlying mock signal  $s_i^*$ .

In this fashion we add the correct power to the Wiener mean reconstruction. The effect of adding the fluctuation term to the Wiener mean is presented in figure 2, where we see the Wiener mean reconstruction, the fluctuation term and the sum of both.

The mock data  $d_i^*$  is generated to obey the data model described in equation (24) and the Wiener variance.

We therefore first draw a mock signal  $s_i^*$  with correct statistics from the multivariate Gaussian signal prior given in equation (9). Such a mock signal is best generated in Fourier space following the description of Martel (2005). One first draws two Gaussian random numbers,  $\chi_a$  and  $\chi_b$ , with zero mean and unit variance and then calculates the real and imaginary part of the signal in Fourier space as:

$$\begin{aligned} RE(\hat{s}_k) &= \sqrt{\frac{P_k}{2}} \chi_a \\ IM(\hat{s}_k) &= \sqrt{\frac{P_k}{2}} \chi_b, \end{aligned} \quad (37)$$

where  $P_k$  is the power-spectrum coefficient at the  $k$ th position in Fourier space. Note, that the mock signal  $s_i^*$  is supposed to be a real quantity, and therefore hermiticity has to be imposed in Fourier space before performing the inverse Fourier transform (for details see Martel 2005).

Next we have to generate the additive noise contribution. In order to draw a noise term with the correct Poissonian statistics, we first draw a random number  $N_i^*$  from the Poissonian distribution:

$$N_i^* \sim \mathcal{P}(N_i^* | \lambda_i^*), \quad (38)$$

where we choose the mean observed galaxy number to be  $\lambda_i^* = n_i^{SF} \bar{\lambda}^2$ . According to equations (18) and (24) the mock noise term  $\epsilon_i^*$  can be calculated as:

$$\epsilon_i^* = \frac{N_i^* - n_i^{SF} \bar{\lambda}^2}{\bar{\lambda}}. \quad (39)$$

It is clear by construction that this mock noise term has vanishing mean and the correct noise covariance matrix. Then, according to equation (24) the mock observation is given as:

$$d_i^* = \sum_k K_{ik} s_k^* + \epsilon_i^*. \quad (40)$$

The proof, that the fluctuation term  $y_i$  as generated by equation (36) truly generates the correct variance is given in Appendix C.

Note, that the application of the Wiener operator is a linear operation, and we can therefore rewrite equation (35) as:

$$s_i = s_i^* + \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} (d_l - d_l^*), \quad (41)$$

where the Wiener operator is applied to the true data  $d_i$  and the mock observation  $d_i^*$  simultaneously. This greatly reduces the CPU time required for the generation of one signal sample.

## 5 SAMPLING THE POWER SPECTRUM

As described above, the signal sampling step provides a noise-less full sky signal sample  $s_i$  consistent with the data. The next step in the Gibbs sampling iteration scheme requires to draw power-spectrum samples from the conditional probability distribution  $\mathcal{P}(S | \{s_i\}, \{d_i\})$ . Since in this Gibbs sampling step the perfect sky signal amplitudes  $s_i$  are known, the power-spectrum is conditionally independent of the data amplitudes  $d_i$ . Hence, in this Gibbs

sampling step, we can sample the power-spectrum from the probability distribution  $\mathcal{P}(S | \{s_i\})$ . In the following we will show that the power-spectrum can easily be drawn from an inverse gamma distribution.

### 5.1 Drawing power-spectrum samples

According to Bayes' theorem, we can rewrite the conditional probability  $\mathcal{P}(S | \{s_i\})$  as:

$$\mathcal{P}(S | \{s_i\}) = \frac{\mathcal{P}(S)}{\mathcal{P}(\{s_i\})} \mathcal{P}(\{s_i\} | S), \quad (42)$$

where  $\mathcal{P}(S)$  is the prior for the signal covariance,  $\mathcal{P}(\{s_i\} | S)$  is given by equation (9) and  $\mathcal{P}(\{s_i\})$  is a normalization constant in this Gibbs sampling step.

More specifically, we are interested in the set of matrix coefficients  $\{S_{ij}\}$  of the covariance matrix  $S$ . As already pointed out in section 4.1 the signal covariance matrix of an homogeneous and isotropic universe, has an especially appealing form in Fourier space, where it takes a diagonal form. In our application the real space covariance matrix coefficients  $\{S_{ij}\}$  are related to their Fourier representation via the fast Fourier transform, as defined in Appendix A. We can therefore write:

$$\begin{aligned} S_{ij} &= C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi i k \frac{\sqrt{-1}}{N}} \hat{S}_{kl} e^{-2\pi i l \frac{\sqrt{-1}}{N}} \\ &= C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi i \frac{\sqrt{-1}}{N} (ik-jl)} \hat{S}_{kl}. \end{aligned} \quad (43)$$

Then we can express the conditional distribution for the Fourier signal covariance coefficients  $\hat{S}_{kl}$  as:

$$\mathcal{P}(\{\hat{S}_{kl}\} | \{s_i\}) = \mathcal{P}(\{S_{ij}\} | \{s_i\}) \left| \frac{\partial \{S_{ij}\}}{\partial \{\hat{S}_{kl}\}} \right|, \quad (44)$$

where

$$\left| \frac{\partial \{S_{ij}\}}{\partial \{\hat{S}_{kl}\}} \right| = \left| \det(\mathcal{J}_{(ij)(kl)}) \right|, \quad (45)$$

is the Jacobian determinant for this coordinate transformation. As the discrete Fourier transform is proportional to a unitary matrix, this Jacobian determinant only amounts to a normalization constant, as has been demonstrated in Appendix B.

With this definition we can rewrite the conditional probability in equation (42), by replacing all the real space covariance matrix coefficients  $S_{ij}$  by their Fourier representation  $\hat{S}_{kl}$ , and normalizing it with the constant obtained from the coordinate transformation. We can therefore write:

$$\begin{aligned} \mathcal{P}(\{\hat{S}_{kl}\} | \{s_i\}) &= \frac{\mathcal{P}(\{\hat{S}_{kl}\})}{C^{N^2} \mathcal{P}(\{s_i\})} \mathcal{P}(\{s_i\} | \{\hat{S}_{kl}\}) \\ &= \frac{\mathcal{P}(\{\hat{S}_{kl}\})}{C^{N^2} \sqrt{\det(2\pi \hat{S})}} \mathcal{P}(\{s_i\}) e^{-\frac{C^2}{2\hat{C}^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} s_k^* \hat{S}_{kl}^{-1} s_l}, \end{aligned} \quad (46)$$

where we used the discrete Fourier transform definition, given in Appendix A, to replace the real space signal amplitudes  $s_i$  by their Fourier counterparts  $\hat{s}_k$ . Introducing equation (10) then allows us to rewrite equation (46) in terms of the power-spectrum coefficients  $P_k$  as:

$$\mathcal{P}(\{P_k\}|\{s_i\}) = \frac{\mathcal{P}(\{P_k\})}{C^{N^2} \mathcal{P}(\{s_i\})} \prod_{k'=0}^{N-1} (2\pi P_{k'})^{-1/2} e^{-\frac{C^2}{2\hat{C}^2} \sum_{k=0}^{N-1} \frac{|s_k|^2}{P_k}}, \quad (47)$$

where the determinant factorizes due to the diagonal form of the signal covariance matrix in Fourier space.

Note, that due to isotropy the power-spectrum is independent of direction in Fourier space, meaning the power-spectrum coefficients only depend on the modulus of the mode vector  $\vec{k}_k$ :

$$P_k = P(\vec{k}_k) = P(|\vec{k}_k|). \quad (48)$$

For this reason, the angular dependence in Fourier space can be summed over.

To do so we remark that the mode vector  $\vec{k}_k$ , as a geometrical object, will not change if we express it in the basis of cartesian coordinates  $\vec{k}_k = k_k^1 \hat{e}_1 + k_k^2 \hat{e}_2 + k_k^3 \hat{e}_3$ , or if we describe it in the basis of spherical coordinates  $\vec{k}_k = k_k^r(\vec{k}_k, \varphi_k, \vartheta_k)$ . We can therefore split the multi index summation into the summation over the three spherical coordinates as:

$$\begin{aligned} \frac{C^2}{\hat{C}^2} \sum_{k=0}^{N-1} \frac{|s_k|^2}{P_k} &= \sum_{|\vec{k}_k|} \frac{1}{P(|\vec{k}_k|)} \sum_{\varphi_k} \sum_{\vartheta_k} \frac{C^2}{\hat{C}^2} \left| \hat{s}(|\vec{k}_k|, \varphi_k, \vartheta_k) \right|^2 \\ &= \sum_{|\vec{k}_k|} \frac{\sigma(|\vec{k}_k|)}{P(|\vec{k}_k|)} \\ &= \sum_{m=0}^{M-1} \frac{\sigma_m}{P_m}, \end{aligned} \quad (49)$$

where we introduced  $\sigma(|\vec{k}_k|) = \sum_{\varphi_k} \sum_{\vartheta_k} C^2 / \hat{C}^2 \left| \hat{s}(|\vec{k}_k|, \varphi_k, \vartheta_k) \right|^2$ , which is the summed signal power on spherical shells around the origin in Fourier space, and the index  $m$  labels each of the  $M$  shells belonging to the different mode vector modulus  $|\vec{k}_k|$  in the Fourier box.

Several different mode vectors  $\vec{k}_k$  may have the same vector modulus  $|\vec{k}_k|$ , and therefore belong to the same shell. To account for this we introduce the number  $n_m$ , which counts the number of different mode vectors  $\vec{k}_k$ , belonging to the  $m$ th shell in Fourier space. This number  $n_m$ , therefore counts the degrees of freedom for each of the  $M$  modes. We can then express the product in equation (47) in terms of  $m$  as:

$$\prod_{k=0}^{N-1} (2\pi P_k)^{-1/2} = \prod_{m=0}^{M-1} (2\pi P_m)^{-n_m/2}. \quad (50)$$

With these definitions equation (47) turns to:

$$\mathcal{P}(\{P_k\}|\{s_i\}) = \frac{\mathcal{P}(\{P_k\})}{C^{N^2} \mathcal{P}(\{s_i\})} \prod_{m=0}^{M-1} (2\pi P_m)^{-n_m/2} e^{-\frac{1}{2} \frac{\sigma_m}{P_m}}, \quad (51)$$

When ignoring the power-spectrum prior  $\mathcal{P}(\{P_k\})$  in the above equation (51), we see that the probability distribution factorizes in the different  $P_m$ , meaning they could be sampled independently.

If also the prior for the different  $P_m$  would factorize as:

$$\mathcal{P}(\{P_k\}) = \prod_{m=0}^{M-1} \mathcal{P}(P_m), \quad (52)$$

then it is possible to sample each mode of the power-spectrum independently.

On large scales, or in the linear regime, the theory of gravitational structure formation tells us that the different Fourier modes evolve independent of each other. In these regimes the proposed power-spectrum prior would be the adequate choice. However, we also know that nonlinear structure formation couples the different

Fourier modes, the stronger the deeper we reach into the nonlinear regime. In these regimes another prior would be more adequate, but also harder to sample.

Anyhow, as already described in section 3 the entire power-spectrum sampling method requires two steps. While the different power-spectrum modes are assumed to be independent in the power-spectrum sampling step, they are not in the signal sampling step. There the different modes are coupled via the observation mask and selection function, and furthermore, the physical coupling of the different modes is represented in the data.

Therefore, in the following we assume a power-spectrum prior, as proposed in equation (52), and defer a more thorough investigation of adequate prior choices in the nonlinear regime to future work.

With this prior choice each mode can be sampled independently from the following probability density distribution:

$$\mathcal{P}(P_m|\{s_i\}) = \frac{\mathcal{P}(P_m)}{(C^{N^2} \mathcal{P}(\{s_i\}))^{1/M}} (2\pi P_m)^{-n_m/2} e^{-\frac{1}{2} \frac{\sigma_m}{P_m}}. \quad (53)$$

Further, we will assume a power-law behavior for the individual mode prior  $\mathcal{P}(P_m) \propto P_m^{-\alpha}$  where  $\alpha$  is a power law index. Note, that a power-law index  $\alpha = 0$  describes the flat prior, while  $\alpha = 1$  amounts to the Jeffrey's prior. The Jeffrey's prior is a solution to a measure invariant scale transformation of the form  $\mathcal{P}(P_m) dP_m = \mathcal{P}(\gamma P_m) \gamma dP_m$  (Wandelt 2004), and therefore is a scale independent prior, as different scales have the same probability.

Inserting this power law prior in equation (53) and imposing the correct normalization, reveals that the power-spectrum coefficients have to be sampled from an inverse gamma distribution given as:

$$\mathcal{P}(P_m|\{s_i\}) = \frac{\left(\frac{\sigma_m}{2}\right)^{(\alpha-1)+n_m/2}}{\Gamma\left((\alpha-1) + \frac{n_m}{2}\right)} \frac{1}{P_m^{\alpha+n_m/2}} e^{-\frac{1}{2} \frac{\sigma_m}{P_m}}. \quad (54)$$

By introducing the new variable  $x_m = \sigma_m / P_m$  and performing the change of variables we yield the  $\chi^2$ -distribution as:

$$\mathcal{P}(x_m|\{s_i\}) = \frac{x_m^{\beta_m/2-1}}{\Gamma(\beta_m/2) (2)^{\beta_m/2}} e^{-\frac{x_m}{2}}, \quad (55)$$

where  $\beta_m = 2(\alpha + n_m/2 - 1)$ . Sampling the power-spectrum coefficients is now an easy task, as it reduces to drawing random samples from the  $\chi^2$ -distribution. A random sample from the  $\chi^2$ -distribution for an integer  $\beta_m$  can be drawn as follows.

Let  $z_j$  be  $\beta_m$  independent, normally distributed random variates with zero mean and unit variance then:

$$x_m = \sum_{j=1}^{\beta_m} z_j^2 = |\vec{z}_m|^2 \quad (56)$$

is  $\chi^2$ -distributed, and  $\vec{z}_m$  is a  $\beta_m$  element vector, with each element being normally distributed. The power-spectrum coefficient sample is then obtained by:

$$P_m = \frac{\sigma_m}{|\vec{z}_m|^2}. \quad (57)$$

It is easy to see that each spectrum coefficient sample is a positive quantity, this ensures that the signal covariance matrix is positive definite as it has to be by definition.

To summarize, we provide an optimal estimator for the power-spectrum coefficients, and their uncertainties.

It is also worth mentioning, that the inverse gamma distribution is a highly non-Gaussian distribution, and that for this reason,

the joint estimates of signal amplitudes  $s_i$  and power-spectrum coefficients  $P_m$  are drawn from a non-Gaussian distribution.

## 5.2 Blackwell-Rao estimator

As described in the introduction, we seek to estimate the probability distribution  $\mathcal{P}(\{P_m\}|\{d_i\})$ , which we can now simply obtain by marginalizing over the signal samples:

$$\begin{aligned} \mathcal{P}(\{P_m\}|\{d_i\}) &= \int d\{s_i\} \mathcal{P}(\{P_m\}|\{s_i\}, \{d_i\}) \mathcal{P}(\{s_i\}|\{d_i\}) \\ &= \int d\{s_i\} \mathcal{P}(\{P_m\}|\{s_i\}) \mathcal{P}(\{s_i\}|\{d_i\}) \\ &\approx \frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} \mathcal{P}(\{P_m\}|\{s_i^j\}), \end{aligned} \quad (58)$$

where  $\{s_i\}^j$  are the signal Gibbs samples, and  $N_{Gibbs}$  is the total number of Gibbs samples.

This result is known as the Blackwell-Rao estimator of  $\mathcal{P}(\{P_m\}|\{d_i\})$  which is guaranteed to have a lower variance than a binned estimator (Wandelt 2004).

It is worth noting, that  $\mathcal{P}(\{P_m\}|\{s_i\})$  has a very simple analytic form, and therefore equation (62) provides an analytic approximation to  $\mathcal{P}(\{P_m\}|\{d_i\})$  based on the Gibbs samples. All the information on  $\mathcal{P}(\{P_m\}|\{d_i\})$  is therefore contained in the  $\sigma_m$  of the individual Gibbs steps, which generate a data set of size  $\mathcal{O}(m_{max} N_{Gibbs})$ , where  $m_{max}$  is the maximal number of independent modes. In addition, to being a faithful representation of  $\mathcal{P}(\{P_m\}|\{d_i\})$  the Blackwell-Rao estimator is also a computationally efficient representation, which allows to calculate any moment of  $\mathcal{P}(\{P_m\}|\{d_i\})$  as:

$$\langle P_m P_{m'} \dots P_{m''} \rangle_{\mathcal{P}(\{P_m\}|\{d_i\})} \approx \frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} \langle P_m P_{m'} \dots P_{m''} \rangle_{\mathcal{P}(\{P_m\}|\{s_i^j\})}, \quad (59)$$

where each of the terms on the right handside can be calculated analytically.

For the inverse gamma distribution given in equation (54) we can then simply calculate the mean of the probability distribution  $\mathcal{P}(P_m|\{d_i\})$  as:

$$\langle P_m \rangle_{\mathcal{P}(P_m|\{d_i\})} \approx \frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} \frac{\sigma_m^j}{2(\alpha - 2) + n_m}, \quad (60)$$

and in analogy the variance as:

$$\langle [P_m - \langle P_m \rangle]^2 \rangle_{\mathcal{P}(P_m|\{d_i\})} \approx \frac{\frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} (\sigma_m^j)^2}{4((\alpha - 2) + n_m/2)^2((\alpha - 3) + n_m/2)}. \quad (61)$$

The Blackwell-Rao estimator also allows us to demonstrate another remarkable property of the Gibbs sampling approach. Although a specific power-spectrum prior has to be employed during the Gibbs analysis of the data, a post processing analysis of the power-spectrum can be performed with any desired power-spectrum prior. Lets assume one prefers to perform a post processing analysis with a power-spectrum prior  $\mathcal{P}'(\{P_m\})$ , rather than with the prior  $\mathcal{P}(\{P_m\})$ , which was employed during Gibbs sampling. We therefore want to estimate the power-spectrum from the following posterior:

$$\begin{aligned} \mathcal{P}'(\{P_m\}|\{d_i\}) &= \mathcal{P}'(\{P_m\}) \frac{\mathcal{P}(\{d_i\}|\{P_m\})}{\mathcal{P}(\{d_i\})} \\ &= \frac{\mathcal{P}'(\{P_m\})}{\mathcal{P}(\{P_m\})} \mathcal{P}(\{P_m\}|\{d_i\}) \end{aligned}$$

$$\begin{aligned} &= \frac{\mathcal{P}'(\{P_m\})}{\mathcal{P}(\{P_m\})} \int d\{s_i\} \mathcal{P}(\{P_m\}|\{s_i\}) \mathcal{P}(\{s_i\}|\{d_i\}) \\ &= \int d\{s_i\} \mathcal{P}'(\{P_m\}) \frac{\mathcal{P}(\{s_i\}|\{P_m\})}{\mathcal{P}(\{s_i\})} \mathcal{P}(\{s_i\}|\{d_i\}) \\ &\approx \frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} \mathcal{P}'(\{P_m\}) \frac{\mathcal{P}(\{s_i^j\}|\{P_m\})}{\mathcal{P}(\{s_i^j\})}, \end{aligned} \quad (62)$$

where we simply made use of the Bayes theorem. Since  $\mathcal{P}(\{s_i\}|\{P_m\})$  is a simple Gaussian distribution, and therefore given analytically, the posterior  $\mathcal{P}'(\{P_m\}|\{d_i\})$  can be calculated with any desired power-spectrum prior in a post-processing step.

## 6 THE PRIOR AND THE COSMIC VARIANCE

The Gibbs sampling procedure consists of the two basic steps, of first sampling perfect noise-less full sky signal samples  $s_i$  and then sampling the power-spectrum coefficients  $P_m$  given  $s_i$ .

Therefore, the probability distribution  $\mathcal{P}(P_m|\{s_i\})$ , given in equation (54), encodes our knowledge on the power-spectrum coefficients  $P_m$ , if we had perfect knowledge of the true signal amplitudes  $s_i$ .

It is clear, that in the case of perfect observations the full posterior distribution for the power-spectrum coefficients  $\mathcal{P}(P_m|\{d_i\})$  would reduce to that one given in equation (54).

This is, because in the case of perfect full-sky and noise-less observations, the signal posterior would collapse to a Dirac delta distribution, due to the vanishing noise covariance matrix. This means, that in this case the signal amplitudes  $s_i$  can be estimated with zero variance.

However, measuring the  $P_m$  to arbitrary precision will never be possible. The power-spectrum coefficients depend on the data through the  $\sigma_m$ , which measure the actual fluctuation power in the observed Universe. It is clear that the probability distribution function (54) for the  $P_m$  will not reduce to a Dirac delta distribution, even though the  $\sigma_m$  have been measured perfectly.

Owing to this fact, there will always remain some uncertainty in the power-spectrum estimation, even in the case of perfect measurements. This residual uncertainty is well known as cosmic variance, which is the direct consequence of only observing just one specific matter field realization.

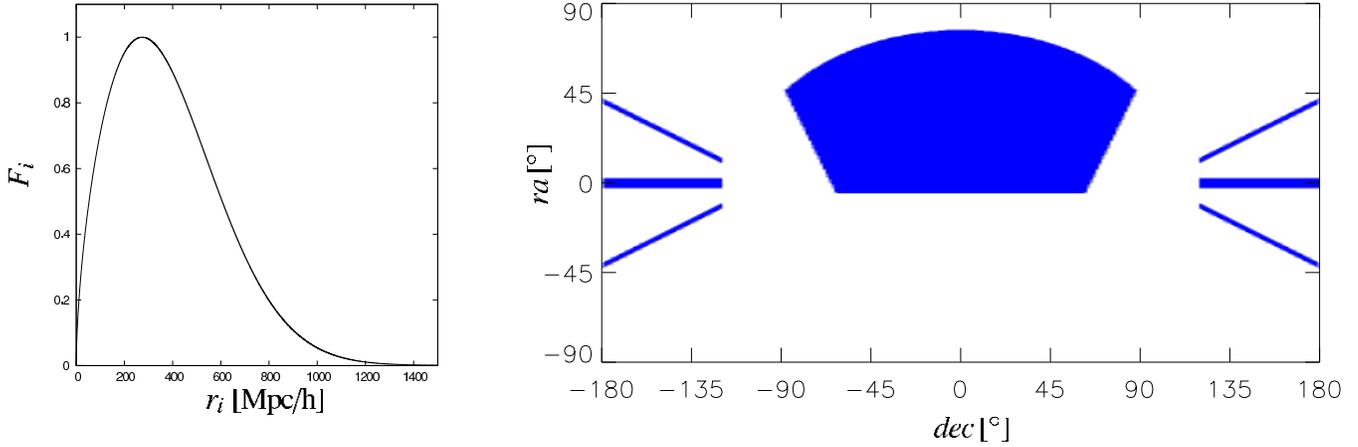
The Gibbs sampling approach, as proposed here, takes this cosmic variance into account, by drawing samples from the probability distribution  $\mathcal{P}(P_m|\{s_i\})$ , which obey the correct statistical properties.

### 6.1 Flat versus Jeffrey's prior

The main characteristics of  $\mathcal{P}(P_m|\{s_i\})$  can be summarized in terms of the mean, mode and variance, which allows us to discuss the influence of the actual power law prior choice. The mode of the inverse gamma distribution (54) is the most frequently used estimator for the power-spectrum coefficients when using fast Fourier transform techniques. The mode  $P_m^*$  is defined as  $P_m^* \in \{P_m | \forall P_l : \mathcal{P}(P_l|\{s_i\}) \leq \mathcal{P}(P_m|\{s_i\})\}$ , which is simply the value of  $P_m$  which maximizes the distribution. For the inverse gamma distribution it is given as:

$$P_m^* = \frac{\sigma_m}{(2\alpha + n_m)}. \quad (63)$$

Assuming a flat prior  $\alpha = 0$ , this immediately returns the frequently used and simple power-spectrum estimator:



**Figure 3.** Selection function and two dimensional sky mask.

$$P_m^* = \frac{\sigma_m}{n_m}, \quad (64)$$

(see e.g. Cui et al. 2008).

However, note, that a flat prior is a questionable choice, when measuring a variance, which is a scale parameter, as it does not correspond to maximal ignorance but biases towards large excursions from zero (Wandelt 2004). Additionally the flat prior does not permit to sample every mode in the three dimensional Fourier box with finite variance, which can be easily seen by looking at the variance of the inverse gamma distribution given as:

$$\langle [P_m - \langle P_m \rangle]^2 \rangle = \frac{\sigma_m^2}{4((\alpha - 2) + n_m/2)^2((\alpha - 3) + n_m/2)} \quad (65)$$

which is only finite for  $2\alpha + n_m > 6$ .

In a three dimensional cubic Fourier box the minimal number  $n_m$  for a mode is  $\min(n_m) = 6$  (except for the zero mode). This corresponds to the six mode vectors  $\vec{k}_k$  with same vector modulus  $|\vec{k}_k|$  along the three axes in Fourier space. Nevertheless, a flat prior ( $\alpha = 0$ ) requires  $n_m > 6$  in order to sample the modes with finite variance, which cannot be fulfilled for these modes.

Therefore, we favor the Jeffrey's prior with  $\alpha = 1$ , which requires only  $n_m > 4$  to sample each mode, except for the zero mode, with finite variance. Jeffrey's prior is also scale invariant, and therefore does not introduce any bias on a log-scale.

## 6.2 Informative prior

The prior discussed in the previous section is a maximal ignorance prior in the sense, that every scale has the same probability. This prior therefore allows for large excursions around the true value of the power-spectrum. This is especially important when sampling the largest scales in a galaxy survey, which are poorly constrained by measurements. A maximum ignorance prior will therefore require to sample a huge space of possible power-spectrum configurations.

However, one can argue, that knowledge about the largest scales exists, either through theory, or CMB measurements, which provide detailed information on the largest scales.

For this reason, it might be interesting to incorporate this a priori knowledge on the power-spectrum into our sampling scheme, and therefore allowing for a more efficient strategy to sample the space of possible power-spectrum configurations.

The most simple informative prior can be obtained by limiting

the range of the Jeffrey's prior, by setting the Jeffrey's prior equal to zero for power-spectrum excursion of more than a certain factor:

$$\mathcal{P}(P_m) \propto \begin{cases} P_m^{-\alpha} & \text{for } P_m^{\text{Prior}}/\tau \leq P_m \leq P_m^{\text{Prior}} \tau \\ 0 & \text{otherwise} \end{cases} \quad (66)$$

where  $P_m^{\text{Prior}}$  is our best guess power-spectrum prior, and  $\tau$  is a factor, which permits a certain range around the prior power-spectrum. The sampling scheme, which arises by implementing this prior, is basically the same as described in section 5.1, with the only exception that power spectrum coefficients  $P_m$ , which do not fulfill the requirement described in equation (66), are rejected and have to be resampled. This prior would still be a maximum ignorance prior over the allowed range.

Another possible informative prior, which allows to sample the entire range, but favoring the region, in which we expect the true power-spectrum to exist, can be very easily found by assuming some a priori knowledge on the  $\sigma_m$ . For example, this can be achieved by incorporating an independent observation to the sampling scheme. In this case we can again assume an inverse gamma distribution for the power-spectrum prior:

$$\mathcal{P}(P_m) = \frac{\left(\frac{\sigma_m^{\text{Prior}}}{2}\right)^{(\alpha^{\text{Prior}}-1)+n_m^{\text{Prior}}/2}}{\Gamma\left((\alpha^{\text{Prior}}-1) + \frac{n_m^{\text{Prior}}}{2}\right)} \frac{1}{P_m^{(\alpha^{\text{Prior}}+n_m^{\text{Prior}}/2)}} e^{-\frac{1}{2} \frac{\sigma_m^{\text{Prior}}}{P_m}}, \quad (67)$$

where  $\sigma_m^{\text{Prior}}$  describes our a priori knowledge on the  $\sigma_m$ ,  $\alpha^{\text{Prior}}$  is the spectral index of our power law prior, which we choose  $\alpha^{\text{Prior}} = 1$  to be the Jeffrey's prior, and  $n_m^{\text{Prior}}$  is the number of mode counts for our prior guess. Note, that the combination of  $\alpha^{\text{Prior}}$  and  $n_m^{\text{Prior}}$  defines how sharp this prior would be. As we want our prior to contain as little information as possible, we choose  $n_m^{\text{Prior}} = 5$  as this is the minimal number of modes, which lead to a finite variance with the Jeffrey's prior.

Introducing an inverse gamma prior will then yield again an inverse gamma distribution for the power-spectrum sampling procedure:

$$\mathcal{P}(P_m | \{s_i\}) = \frac{\left(\frac{\sigma_m^{\text{Prior}} + \sigma_m}{2}\right)^{(\alpha^{\text{Prior}}-1)+(n_m^{\text{Prior}}+n_m)/2}}{\Gamma\left((\alpha^{\text{Prior}}-1) + \frac{(n_m^{\text{Prior}}+n_m)}{2}\right)} \frac{e^{-\frac{1}{2} \frac{(\sigma_m^{\text{Prior}} + \sigma_m)}{P_m}}}{P_m^{(\alpha^{\text{Prior}}+(n_m^{\text{Prior}}+n_m)/2)}}. \quad (68)$$

By introducing  $x_m = (\sigma_m^{\text{Prior}} + \sigma_m)/P_m$ , this can again be rewritten as a  $\chi^2$ -distribution:

$$\mathcal{P}(x_m|\{s_i\}) = \frac{x_m^{\beta_m/2-1}}{\Gamma(\beta_m/2) (2)^{\beta_m/2}} e^{-\frac{x_m}{2}}, \quad (69)$$

where  $\beta_m = 2(\alpha^{\text{Prior}} + (n_m^{\text{Prior}} + n_m)/2 - 1)$ .

A power-spectrum sample is then obtained in the same fashion as described in section 5.1, by

$$P_m = \frac{\sigma_m^{\text{Prior}} + \sigma_m}{|\vec{z}_m|^2}, \quad (70)$$

with  $\vec{z}_m$  being  $\beta_m$  element vector, with each element being normally distributed with zero mean and unit variance.

As an example of incorporating theoretical information one can generate the  $\sigma_m^{\text{Prior}}$  by:

$$\sigma_m^{\text{Prior}} = n_m^{\text{Prior}} P_m, \quad (71)$$

which will yield on average the prior power-spectrum. In this fashion the Gibbs sampler will sample around our prior guess for the power-spectrum.

Note, that this method provides a possible interface for joint power-spectrum estimation from a joint CMB and large scale structure analysis, where the  $\sigma_m^{\text{Prior}} = \sigma_m^{\text{CMB}}$  will be obtained from a CMB analysis step.

### 6.3 Hidden prior

In the previous section, we described how to sample each mode  $|\vec{k}_k|$  of the Fourier box individually, which yields extremely fine frequency resolution in the power-spectrum estimate. As in practice such a high frequency resolution is not required, or desired, one allows each shell  $m'$  to have a finite thickness  $\Delta k_{m'}$ , rather than treating it as infinitely thin.

As the newly designated shells  $m'$  have a finite thickness, different infinitely thin shells  $m$  now belong to the same shell  $m'$ .

With the shells having a finite thickness, different close by Fourier modes  $|\vec{k}_k|$  fall into the same bin, and therefore an assumption about the functional shape  $f_m(|\vec{k}_k|)$ , which yields the correct weighting, over this shell  $m'$  around the central mode  $|\vec{k}_{m'}$  must be assumed:

$$P_m = A_{m'} f_m^{m'} \text{ for } \vec{k}_k \in [|\vec{k}_m - \Delta k_m/2, |\vec{k}_m + \Delta k_m/2], \quad (72)$$

where  $A_{m'}$  is the constant amplitude for the  $m'$ th shell. Usually, the shape of the power-spectrum  $f_m^{m'}$  is assumed to be constant over the shell width  $\Delta k_{m'}$ , which amounts to "binning" the power-spectrum. However, in general  $f_m^{m'}$  could assume any desired functional shape.

With this assumption, the estimate of the power-spectrum is further constrained by the assumed functional shape  $f_m^{m'}$ , and we seek to estimate the set of power-spectrum coefficients  $\{P_m\}$  from a probability distribution given as:

$$\begin{aligned} \mathcal{P}(\{P_m\}|\{s_i\}, \{f_m^{m'}\}) &= \frac{\mathcal{P}(\{f_m^{m'}\}|\{P_m\})}{\mathcal{P}(\{f_m^{m'}\}|\{s_i\})} \frac{\mathcal{P}(\{P_m\})\mathcal{P}(\{s_i\}|\{P_m\}, \{f_m^{m'}\})}{\mathcal{P}(\{s_i\})} \\ &= \frac{\mathcal{P}(\{f_m^{m'}\}|\{P_m\})}{\mathcal{P}(\{f_m^{m'}\}|\{s_i\})} \frac{\mathcal{P}(\{P_m\})\mathcal{P}(\{s_i\}|\{P_m\})}{\mathcal{P}(\{s_i\})}, \end{aligned} \quad (73)$$

where we assumed conditional independence  $\mathcal{P}(\{s_i\}|\{P_m\}, \{f_m^{m'}\}) = \mathcal{P}(\{s_i\}|\{P_m\})$  of the functional shape, once all power-spectrum coefficients are given.

The usual implicit assumption when introducing this kind of power-spectrum binning is:

$$\mathcal{P}(\{f_m^{m'}\}|\{P_m\}) \propto \prod_m \delta^D(A_{m'} f_m^{m'} - P_m), \quad (74)$$

meaning, we claim to know the exact functional shape of the power-spectrum within the shells  $m'$ .

This reduces the amount of free parameters to be sampled, from the set of  $N_m$  power-spectrum coefficients  $P_m$  to the set of  $N_{m'}$  amplitudes  $A_{m'}$ .

If, for instance, we knew the exact shape of the entire power-spectrum, sampling the power-spectrum could be reduced to the task of just sampling the overall amplitude. Such an approach is pursued, when trying to sample the cosmological parameters.

However, with the above definition, sampling the power-spectrum reduces to sample the amplitudes  $A_{m'}$ , and we can write:

$$\begin{aligned} \mathcal{P}(\{A_{m'}\}|\{s_i\}, \{f_m^{m'}\}) &= \mathcal{P}(\{P_m\}|\{s_i\}, \{f_m^{m'}\}) \left| \frac{\partial\{P_m\}}{\partial\{A_{m'}\}} \right| \\ &= \prod_{m'} \frac{\left(\frac{\sigma_m}{2 f_m^{m'}}\right)^{(\alpha-1)+n_m/2}}{\Gamma((\alpha-1) + \frac{n_m}{2})} \frac{e^{-\frac{1}{2} \frac{\sigma_m}{A_{m'} f_m^{m'}}}}{(A_{m'})^{(\alpha+n_m/2)}} \\ &= \prod_{m'} \prod_{m \in m'} \frac{\left(\frac{\sigma_m}{2 f_m^{m'}}\right)^{(\alpha-1)+n_m/2}}{\Gamma((\alpha-1) + \frac{n_m}{2})} \frac{e^{-\frac{1}{2} \frac{\sigma_m}{A_{m'} f_m^{m'}}}}{(A_{m'})^{(\alpha+n_m/2)}} \\ &= \prod_{m'} \frac{\left(\frac{\sigma_{m'}}{2}\right)^{(\alpha-1)+n_{m'}/2}}{\Gamma((\alpha-1) + \frac{n_{m'}}{2})} \frac{e^{-\frac{1}{2 A_{m'}} \sum_{m \in m'} \frac{\sigma_m}{f_m^{m'}}}}{(A_{m'})^{(\alpha+n_{m'}/2)}} \\ &= \prod_{m'} \frac{\left(\frac{\sigma_{m'}}{2}\right)^{(\alpha-1)+n_{m'}/2}}{\Gamma((\alpha-1) + \frac{n_{m'}}{2})} \frac{e^{-\frac{1}{2} \frac{\sigma_{m'}}{A_{m'}}}}{(A_{m'})^{(\alpha+n_{m'}/2)}}, \end{aligned} \quad (75)$$

where  $\sigma_{m'} = \sum_{m \in m'} \sigma_m / f_m^{m'}$  and  $n_{m'} = \sum_{m \in m'} n_m$ . As this probability distribution factorizes in the amplitudes  $A_{m'}$ , each of these amplitudes can be independently sampled from the inverse gamma distribution, with the method described in section 5.1.

A power-spectrum coefficient  $P_m$  is therefore obtained as:

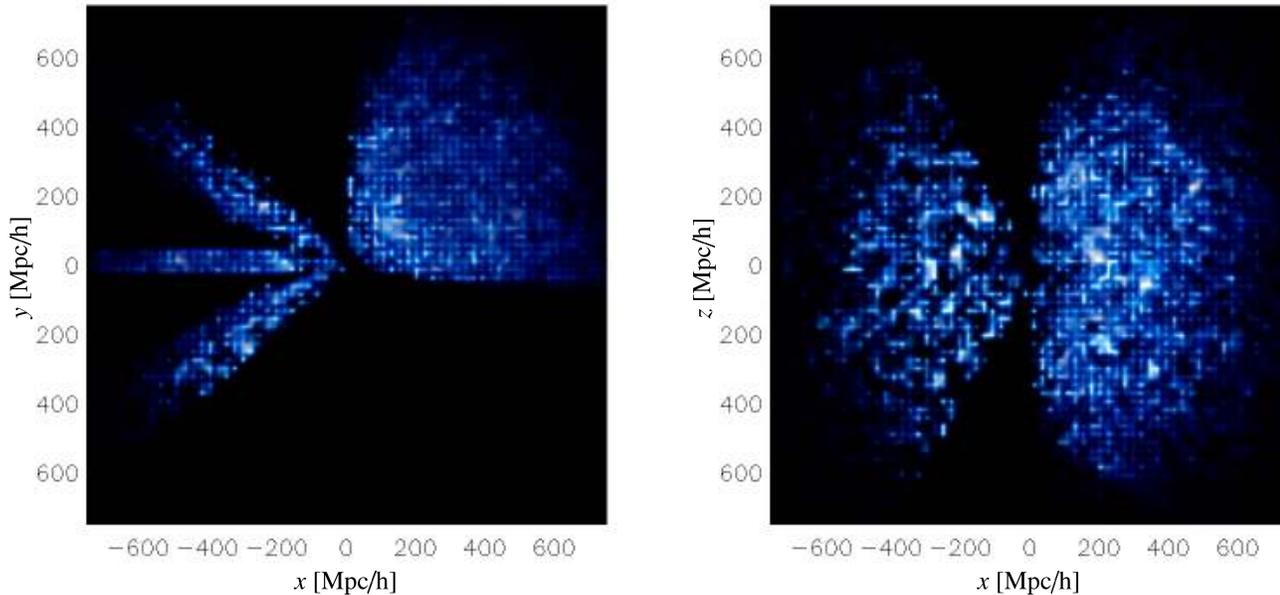
$$P_m = \frac{\sigma_{m'} f_m^{m'}}{|\vec{z}_{m'}|^2}, \quad (76)$$

where  $\vec{z}_{m'}$  is a  $n_{m'}$  component vector, with the elements being Gaussian distributed with zero mean and unit variance.

Note, since  $n_{m'} \geq n_m$  the variance for the power-spectrum coefficients  $P_m$  is reduced. This is the result of reducing the amount of free parameters, by introducing "binning", as now each amplitude estimate for the  $A_{m'}$  is based on more supporting points than the individual  $P_m$ .

Due to the finite shell width  $\Delta k_{m'}$  neighboring modes are coupled. This fact could be exploited to circumvent the problem of missing mode coupling in the nonlinear regime. For example, if the different shells would be logarithmically spaced, it is possible to sample the largest scales independently, while towards the nonlinear regime, the modes get more and more coupled. From a physical point of view, the logarithmic spacing would therefore be best suited for this problem. On the other hand, introducing "binning" to the power-spectrum sampling procedure, makes the method insensitive to fluctuations on scales smaller than the shell width  $\Delta k_{m'}$ , and a  $\Delta k_{m'}$  should therefore be chosen carefully in order not to miss features we intend to recover.

It is also important to note, that the variance in the power-spectrum sampling step defines the stepsize for the random walk, to sample the joint probability distribution of signal and power-spectrum. If the "binning" is unreasonable large, and therefore variance is dramatically reduced, it takes much longer to explore the joint space of signal amplitudes  $s_i$  and power-spectrum coefficients



**Figure 4.** Volume rendering of observed galaxy densities in two different projections.

$P_m$ . For this reason, we prefer to sample with rather high spectral resolution for the power-spectrum.

## 7 NUMERICAL IMPLEMENTATION

Our numerical implementation of the large scale structure Gibbs sampler is called ARES (Algorithm for REconstruction and Sampling). It can be separated into the described two Gibbs sampling steps of estimating a signal sample, which involves solving large systems of equations, and sampling the power-spectrum, by drawing random samples from the inverse gamma distribution.

### 7.1 Inversion of matrices

The signal sampling step is by far the most numerically demanding step as it requires fast and efficient inversions of large matrices. ARES utilizes the fast operator based conjugate gradients inversion technique as presented in the ARGO code (Kitaura & Enßlin 2008), which has recently been applied to Sloan Digital Sky Data, to obtain matter field reconstructions (Kitaura et al. 2009). Operator based inversion techniques have been previously developed for CMB data analysis (Wandelt 2004; Eriksen et al. 2004; Jewell et al. 2004).

Rather than requiring to store the matrices under consideration explicitly in computer memory, which would be numerically prohibitive, this approach only requires to know how these matrices would act on a vector, and therefore reduces the required amount of computer memory to numerically feasible amounts. Further, it is possible to reduce the scaling of the most expensive matrix operation to that one of a fast Fourier transform. ARES uses the FFTW3 library, which therefore reduces the scaling of the most expensive operation to  $O(N \log(N))$  (Frigo & Johnson 2005).

The FFTW3 library also incorporates the feature of parallel Fourier transforms, which allows for straight forward parallelization of our code.

### 7.2 Random number generation

Our random number generation relies on a pseudo random number generator as provided by the GNU scientific library (gsl) (Galassi et al. 2003). In particular, we use the Mersenne Twister MT19937, with 32-bit word length, as provided by the `gsl_rng_mt19937` routine.

The Mersenne Twister algorithm was designed for Monte Carlo simulations, where primarily good quality numbers and speed are decisive (Matsumoto & Nishimura 1998). It has been proven to have a period of  $2^{19937} - 1$ , where for our application in practice we usually require only about  $2^{35}$  unique random numbers. Also note, that the very high order of dimensional equidistribution guarantees negligible serial correlation in the output sequence. The Mersenne Twister algorithm passed several tests for statistical randomness, including the Diehard tests.

### 7.3 Parallelization

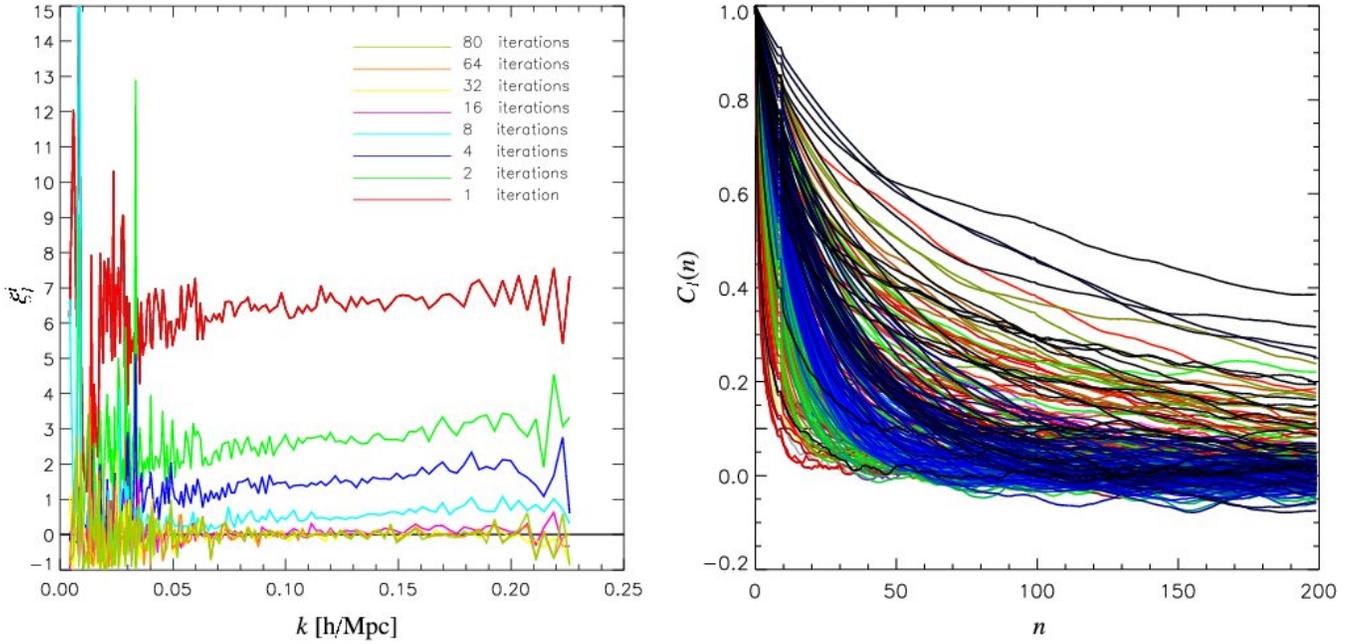
Parallelization of the code is a crucial issue, since CPU time is the main limiting factor of our method. Even though the conjugate gradient method allows for very efficient matrix inversions, for a complete data analysis one has to perform many of these matrix inversions, and therefore the total prefactor of the algorithm increases.

In figure 6 we show a typical progression for the wall clock times of the map making algorithm during the Gibbs sampling chain for a low resolution simulation with  $N_{VOX} = 64^3$  and a setup as described in section 8.

As can be seen, the wall clock times may vary from Gibbs iteration to Gibbs iteration.

This variation is due to complexity of the actual problem in the Gibbs iteration, which may require more conjugate gradients steps to reach the desired numerical accuracy.

The average creation time for one Gibbs sample in this low resolution simulation is about 20 seconds. Where this test was run on a single Desktop CPU. Doubling the resolution will require



**Figure 5.** Test results from low resolution test run. The left panel shows the results of the Burn-in test for eight sequential  $\xi_i^2$  as a function of Fourier modes  $k$ . The right panel displays results for the correlation coefficients  $C_l(n)$  of the correlation length test for all Fourier modes.

roughly eight times longer to create a sample. Hence, a  $N_{\text{vox}} = 128^3$  simulation will require roughly 5 minutes and a  $N_{\text{vox}} = 512^3$  about 8 hours to create a single Gibbs sample.

This immediately clarifies the need to parallelize the code, as usually tenth of thousands of Gibbs samples are required.

There are in principle two different approaches to parallelize the code.

The most demanding step in the Gibbs sampling chain is the map making process. One could therefore parallelize the map making algorithm, which in principle requires to parallelize the fast Fourier transform. The FFTW3 library provides parallelized fast Fourier transform procedures, and implementation of those is straight forward (Frigo & Johnson 2005). However, optimal speed up cannot be achieved. The other approach relies on the fact that our method is a Monte Carlo process, and each CPU can therefore calculate its own Markov chain. In this fashion we gain optimal speed up and the possibility to initialize each chain with different initial guesses.

The major difference between these two parallelization approaches is, that with the first method one tries to calculate a rather long sampling chain, while the latter one produces many shorter chains.

As we will see in the next section, successive Gibbs samples are highly correlated in the low signal to noise regime and producing a larger number of independent samples requires longer chains. This problem, however, can be partially overcome by initializing each Markov chain with an independent power-spectrum guess.

## 8 TESTING ARES

In this section, we apply ARES to simulated mock observations, where the underlying dark matter signal is perfectly known. With these mock observations we will be able to demonstrate that the code produces results consistent with the theoretical expectation. In

addition, we will gain insight in how the code may perform in real-world applications, when CPU time is limited. Therefore, we will set up Gaussian mock cases, designed to highlight some specific feature of the code.

### 8.1 Setting up a Gaussian Mock observation

For the cases studied in this section we set up a set of low resolution mock observations based on Gaussian random fields, which will allow us to calculate a large number of samples in reasonable computational times.

These mock observations are generated on a three dimensional cartesian box with  $N_{\text{side}} = 64$ , corresponding to  $N_{\text{vox}} = 262144$  volume elements, and a box length of  $L = 1500$  Mpc, with the observer positioned at the center.

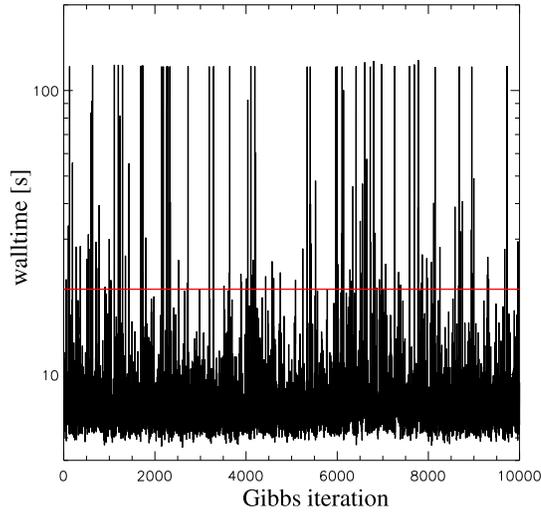
The mock observations are generated according to the procedure described in section 4.3, with the underlying cosmological power-spectrum being calculated, with baryonic wiggles, following the prescription described in Eisenstein & Hu (1998) and Eisenstein & Hu (1999). For these simulations we assumed a standard  $\Lambda$ CDM cosmology with the set of cosmological parameters ( $\Omega_m = 0.24$ ,  $\Omega_\Lambda = 0.76$ ,  $\Omega_b = 0.04$ ,  $h = 0.73$ ,  $\sigma_8 = 0.74$ ,  $n_s = 1$ ).

To generate the uncorrelated noise component, we assume a noise structure function  $n_i^{\text{SF}} = M_i F_i / \bar{\lambda}$ , with  $\bar{\lambda} = 8.0 \times 10^{-3} L^3 / N_{\text{vox}}$  in voxel space. Then we draw random Poisson samples via the procedure described in section 4.3.

Note, that this is equivalent to drawing a galaxy distribution from the ensemble mean dark matter density, which has been obtained by averaging over all possible matter field realizations.

The survey properties are described by the galaxy selection function  $F_i$  and the observation Mask  $M_i$ . The selection function is given by:

$$F_i = \left(\frac{r_i}{r_0}\right)^b \left(\frac{b}{\gamma}\right)^{-b/\gamma} e^{\frac{b}{\gamma} - \left(\frac{r_i}{r_0}\right)^\gamma}, \quad (77)$$



**Figure 6.** Wall clock times for a single map making step over 10000 Gibbs iterations. The red line denotes the average wall clock time for one matrix inversion step.

where  $r_i$  is the comoving distance from the observer to the center of the  $i$ th voxel. For our simulation we chose parameters  $b = 0.6$ ,  $r_0 = 500$  Mpc and  $\gamma = 2$ .

In figure 3 we show the selection function together with the sky mask, which defines the observed regions in the sky. The two dimensional sky mask is given in sky coordinates of right ascension and declination. The projection of this two dimensional mask into the three dimensional volume yields the three dimensional mask  $M_i$ .

Two different projections of this generated mock galaxy survey are presented in figure 4 to give a visual impression of the artificial low resolution observation.

## 8.2 Testing convergence and correlations

The theory of Gibbs sampling states that the individual Gibbs samples converge to being samples from the joint probability distribution. However, the theory itself does not provide any criterion to detect when the samples start being samples from the joint probability distribution. Therefore, this initial burn-in behavior has to be tested by experiments. Further we will follow a similar analysis as described in Eriksen et al. (2004), to estimate the convergence behavior and the correlation length of the Gibbs samples.

This analysis is important, as it allows us to understand the performance of our code in real world applications, which is particularly relevant for estimating how many Gibbs samples are required for an accurate power-spectrum estimation.

We study the initial burn-in behavior by setting up a simple experiment, in which we set the initial guess for the power-spectrum to be ten times larger in amplitude than the true underlying power-spectrum. Then ARES is applied to the mock observation, as described in the previous section 8.1, to calculate a number of Gibbs iterations. The obtained power-spectrum samples  $P_l^i$  of the  $i$ th iteration are then compared to the true power-spectrum via:

$$\xi_l^i = \frac{P_l^i - P_l^{true}}{P_l^{true}}, \quad (78)$$

where the  $P_l^{true}$  are the true power-spectrum coefficients, with which the mock dark matter signal was simulated.

The results for the  $\xi_l^i$  are shown in figure 5. Here, we observe

a systematic drift of the successive power-spectra towards the true underlying power-spectrum. Further we see that the initial burn-in phase, for the kind of setup as demonstrated here, is rather short. The algorithm requires about 30 Gibbs iteration after which we can assume the samples to be samples from the joint probability distribution. Also note, that in this test, we do not observe any particular hysteresis for the poorly constrained large scale modes, meaning they do not remain at their initially set values. This demonstrates the ability of the code to handle correctly the mode coupling introduced by the sky cut.

However, it is clear that a poor initial guess invalidates a certain number of samples, especially at large scales, where the uncertainty due to the sky mask dominates. For this reason, it is advantageous to initialize the Gibbs sampling chain with an initial guess, which is close to the true power-spectrum, to ensure short burn-in times. As can be seen in figure 5, any bad initial guess would reveal itself by a systematic drift in the Gibbs chain, and can therefore be detected easily.

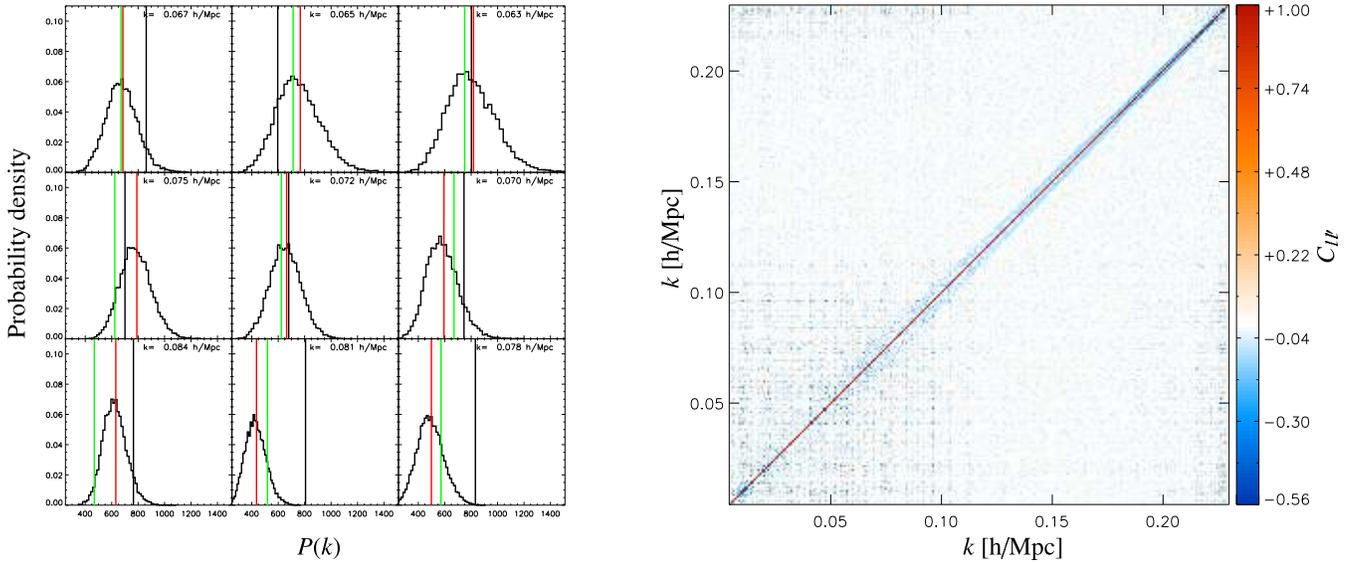
Next, we want to analyze the correlation of the individual Gibbs samples in the sequence. This is a crucial point, as it permits us to estimate the number of independent samples, which can be obtained from a Gibbs chain of given length.

The correlation between sequential Gibbs samples can be best understood by reviewing the sampling algorithm.

A Gibbs sample of the joint probability distribution of signal and power-spectrum is obtained in two steps. In the first step a Wiener reconstruction is performed, based on the assumption of a given power-spectrum, and the power lost due to noise filtering, masks and selection effects is replaced by a fluctuation term. The signal obtained in this first sampling step mimics a full sky noise-less signal. It is clear, that the power-spectrum of this signal is determined by the data in the high signal-to-noise region and by the assumed power-spectrum in the low signal-to-noise region.

In the second step the power-spectrum sample is generated, based on this full sky noise-less signal sample, obtained in the previous signal sampling step. The obtained power-spectrum then works as input power-spectrum to the next Gibbs step, and the iteration starts again.

In this fashion the Gibbs sampler performs a random walk in the multi-dimensional space of signal map and power-spectrum.



**Figure 7.** The left panel shows the mode statistics for some selected Fourier modes. The green vertical line denotes the true underlying power-spectrum. The red line denotes the averaged mean obtained from the Gibbs chain, and the black vertical line denotes the power-spectrum of the specific matter field realization. The right panel shows the power-spectrum covariance matrix estimated from the 40,000 Gibbs samples.

The stepsize of the power-spectrum sampling step is solely determined by cosmic variance, and does not take into account the noise variance, as described in section 5.1.

However, we want to probe the full probability distribution, which includes both noise and cosmic variance. This difference does not matter in the high signal to noise regime, since there cosmic variance will dominate the total variance, and for any practical purposes all Gibbs samples will be uncorrelated in this regime. This, however, is not true in the low signal to noise regime. Since the random stepsize between two subsequent samples is determined only by the cosmic variance, and not by the much larger noise variance, two sequential samples will be strongly correlated. In this case a longer Gibbs sequence is required to produce uncorrelated samples.

Reducing the variance by introducing binning to the power-spectrum, as described in section 6.3, will lead to even longer correlation length in the Gibbs chain. This simply means, the joint probability distribution will be sampled with a finer resolution in power-spectrum space.

We study this correlation effect by assuming the power-spectrum coefficients  $P_l$  of different modes  $l$  in the Gibbs chain to be independent and estimate their correlation in the chain by calculating the autocorrelation function:

$$C_l(n) = \left\langle \frac{P_l^i - \langle P_l \rangle}{\sqrt{\text{Var}P_l}} \frac{P_l^{i+n} - \langle P_l \rangle}{\sqrt{\text{Var}P_l}} \right\rangle, \quad (79)$$

where  $n$  is the distance in the chain measured in iterations (for a similar discussion in case of the CMB see Eriksen et al. 2004).

We can then define the correlation length of the Gibbs sampler as the distance in the chain  $n_C$  beyond which the correlation coefficient  $C_l(n)$  has dropped below 0.1.

The results for the different modes  $l$  are presented in figure 5. As one can see, the vast majority of the different Fourier modes have a correlation length of about  $n_C \sim 100$  Gibbs iterations. The rest of the modes show increasingly longer correlation length, which increases towards the highest frequencies contained in the

box. For this reason, the Nyquist frequency has the longest correlation length.

Especially for the highest frequencies towards the Nyquist frequencies, noise domination is only a partial explanation for the long correlation length. The far more important fact is, that the Nyquist frequency can in general not be resolved properly by the data, for example the Nyquist frequency is contained only once in the Fourier box. For this reason, the variance increases towards the Nyquist frequency. Note, that this effect arises from the technical implementation of the fast Fourier transform, which operates on a finite grid, and that we are in principle able to account for these technical effects of the analysis scheme itself. However, the long correlation length for the Nyquist frequencies will in general only provide a rare amount of independent estimates at the Nyquist frequency and therefore this must be taken into account in further analysis.

It is clear, that this effect shifts to higher frequencies as soon as the resolution of our analysis scheme is increased. This can be observed in the high resolution analysis discussed in the next section 8.3. Also note, that this effect can be cured by introducing an informative prior which will greatly reduce the correlation length at these frequencies, as shown later in section 8.4.

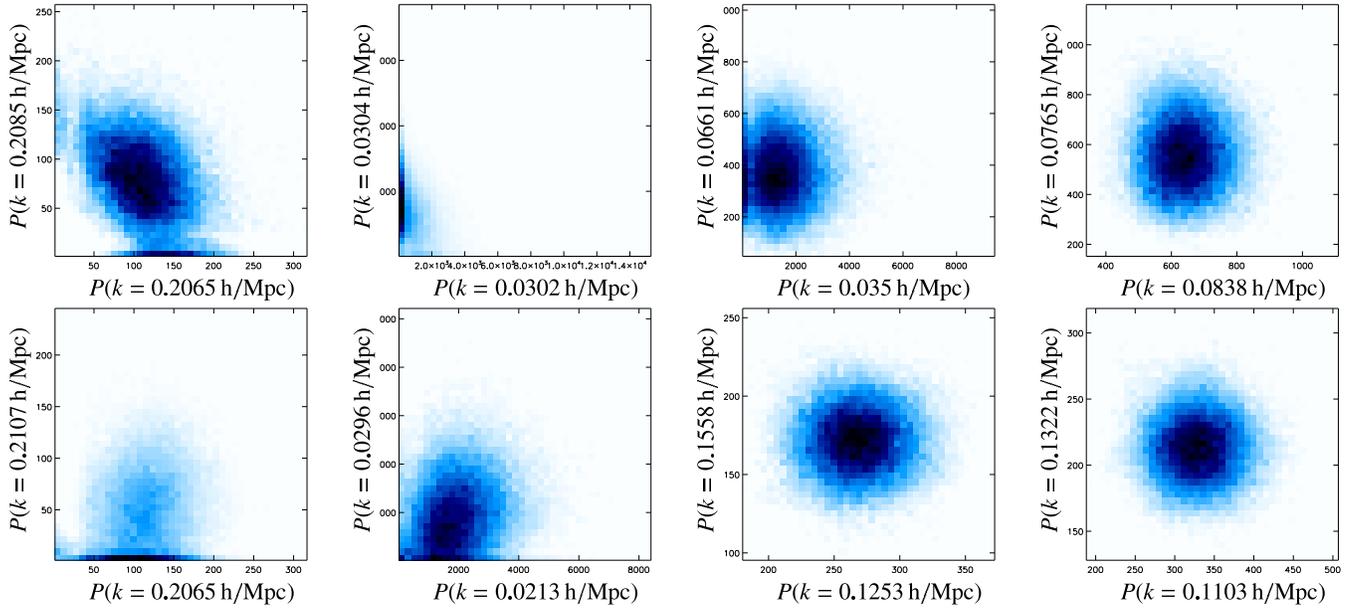
To study the marginalized posterior  $P_l$  distributions in more detail we plot their histograms in figure 7. It is worth mentioning that none of them is even approximately Gaussian.

Another crucial point to address is the question how well we were able to account for effects of the survey geometry. This information is contained in the correlation structure of the estimates. Therefore, we can examine this effect by calculating the correlation matrix of the  $P_l$  estimates:

$$C_{l'l'} = \left\langle \frac{P_l - \langle P_l \rangle}{\sqrt{\text{Var}P_l}} \frac{P_{l'} - \langle P_{l'} \rangle}{\sqrt{\text{Var}P_{l'}}} \right\rangle, \quad (80)$$

where the ensemble averages are taken over 40,000 Gibbs samples.

We present the result in figure 7. It can be clearly seen that this correlation matrix has a very well defined diagonal structure, as expected from theory. The highest off-diagonal correlations have been



**Figure 8.** 2D marginalized posterior densities. Each plot shows the full joint posterior of the data, integrated over all dimensions except for the two shown.

measured to be less than 20%, and are found at the highest frequencies close to the Nyquist frequency. Figure 7 also shows a blue band of anti-correlation around the diagonal. This anti-correlation indicates that the power-spectrum frequency resolution is higher than supported by the data. Since the data is fixed, and the mask couples neighbored Fourier modes, an increase in the power-spectrum amplitude  $P_l$  has to be compensated with a decrease in the neighboring power-spectrum amplitude  $P_{l+1}$  to have a good fit to the data. It is therefore possible, in a post-processing step, to reduce the frequency resolution of the estimated power-spectra until the anti-correlation vanishes. This is the idea behind the running median estimator, which will be presented in section 9.

However, since the posterior distributions for the  $P_l$  are non-Gaussian, the two point correlations do not contain all information. For this reason, we also demonstrate the marginalized posterior distribution for pairs of  $P_l$ s in figure 8, where we also show examples of maximally correlated and maximally anti-correlated modes.

Finally, we have plotted the full spectrum computed from our 40,000 sample run in figure 9. As can be seen, we chose a very high frequency resolution to reduce the correlation length of the Gibbs sampler in the low signal to noise regime. The estimated ensemble mean power-spectrum follows the true underlying power-spectrum, in particular the baryonic features. Towards the large scales, the uncertainty increases, as expected, since due to survey geometry and selection effects these scales are only poorly constrained by the data.

### 8.3 High resolution Simulation

In the previous section, we performed a low resolution analysis to compute a sufficiently large set of samples to estimate the correlation behavior of our algorithm. However, such a large amount of samples is not necessarily required and computational time may be better invested in performing higher resolution analysis. Therefore, in this section we describe the results obtained from such a high resolution analysis.

The setup for this test is the same as described in section 8.1,

with the exception that here we use  $256^3$  voxels on an equidistant grid.

The main limitation for this test is CPU time, as a single sampling step takes about one hour. For this reason, extremely long chains are not feasible. Hence, we run many, rather short, parallel Gibbs chains, as described in section 7.3. In this fashion we obtain an optimal speed up for this parallelization scheme. The Gibbs sampler was run over 24 independently initialized chains and provided a total of 4230 samples.

We present the power-spectra obtained from the multiple-chain analysis in figure 9. As can be seen there is overall good agreement with the realization specific power-spectrum and the ensemble mean estimate found by the Gibbs sampler. We also do not observe a detectable bias in any parts of the spectrum.

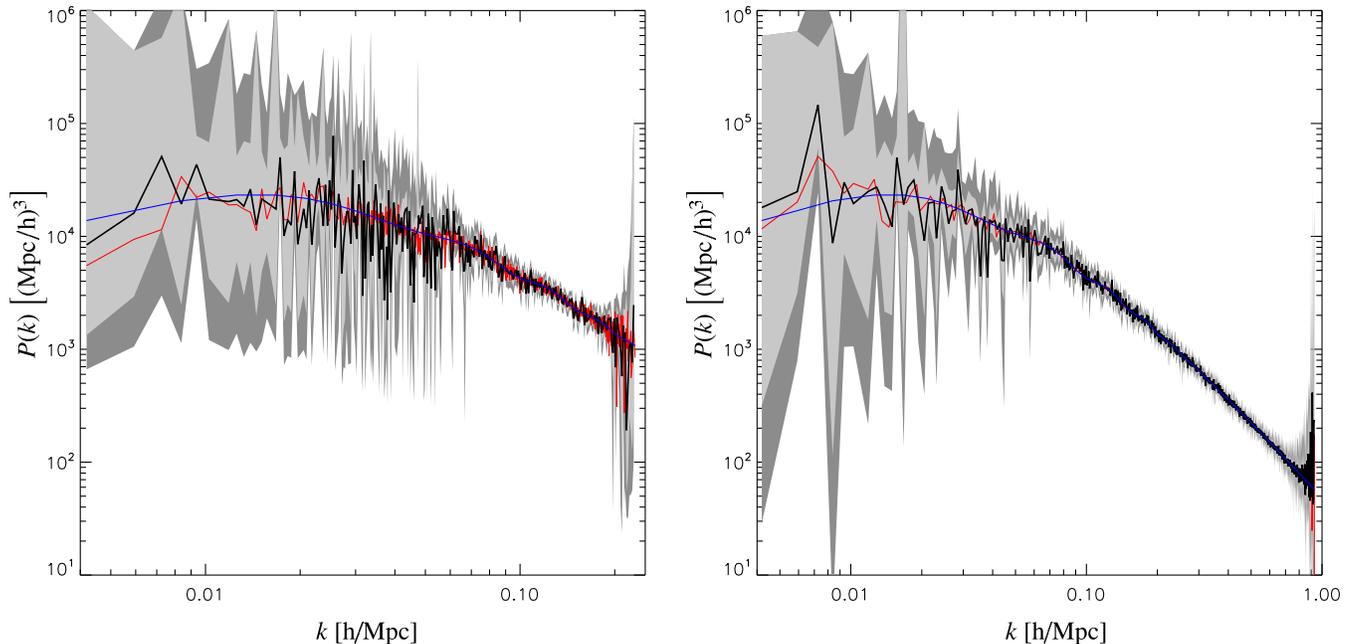
Towards the Nyquist frequency the uncertainty increases. This is expected, as towards the edges of the box, the number of modes decreases, and variance increases. Note, that in this fashion the method takes into account the uncertainty introduced by the analysis scheme itself, for example the Fourier space discretization of the FFT.

### 8.4 Testing an informative Prior

When trying to analyze real galaxy surveys, one is faced with the situation that due to sky cuts and selection effects, usually less than 30% of the volume is available for analysis. This affects mainly the estimate of the largest scales in the survey, as they are sparsely sampled, and therefore poorly constrained by the data.

In this situation the Jeffrey's prior, as a maximal ignorance prior, allows for large excursions from the expected true underlying power-spectrum. Since the Jeffrey's prior provides equal probability for all scales between zero and plus infinity, it also allows for power-spectrum values, which can be excluded on theoretical grounds, or by complementary experiments, which are more sensitive at the largest scales, such as CMB experiments.

From a Bayesian point of view, one could argue, that in the



**Figure 9.** Power-spectrum estimates obtained from the Gibbs sampling chain for a low resolution (left panel) and a high resolution run (right panel). The black lines represent the ensemble mean of the sample set and the light gray and dark gray shaded regions denote the one and two sigma confidence regions respectively. Additionally we show the according input power-spectra. The blue line shows the cosmological power-spectrum from which the matter field realization was drawn, and the red line is the power-spectrum of this specific matter field realization.

presence of a priori knowledge, a maximal ignorance prior is not the optimal choice.

Rather than sampling the entire space of possible power-spectrum coefficients  $P_m$  with equal probability, it would be beneficial to preferably sample the region in which we expect the true power-spectrum to exist and allowing for larger excursions with smaller probability.

This would have the effect, that the region of interest would be sampled more densely, and therefore allowing for better power-spectrum estimates with the same amount of Gibbs samples. Also remember, that according to the discussion in section 5.2, the prior can be changed for any final post-processing analysis.

As an informative prior can lead to a more efficient sampling strategy in the presence of a priori knowledge, in the following we test ARES when employing the inverse gamma prior as described in section 6.2.

We base the inverse gamma prior on a flat power-spectrum guess, which was calculated according to Eisenstein & Hu (1998) and Eisenstein & Hu (1999), without the baryonic wiggles. We explicitly do not incorporate the information on the baryonic features, to demonstrate that solely the data drives their estimate. Further, we choose  $n_m^{prior} = 5$ , in order to make the prior sufficiently broad, while at the same time ensuring that it has finite variance.

With this prior choice we repeat the standard testing procedure as described in section 8.1.

At first we test the initial burn-in time, by starting with a power-spectrum which is a factor 10 higher in amplitude than the underlying true power-spectrum.

The results for the according  $\xi_i^j$ , described in equation (78), are presented in figure 10. It can be seen that the burn-in time is much shorter for the large scale modes, which are poorly constrained by the data. Also note, that the overall burn-in times for the power-spectra are comparable to those of the maximal ignorance prior

case (see figure 5), indicating that these modes are not influenced by the informative inverse gamma prior.

The real advantage of the informative prior becomes obvious when analyzing the correlation length for the individual power-spectrum coefficients  $P_m$  in the Gibbs chain. Again we used a low resolution  $64^3$  Voxel simulation in order to estimate the correlation coefficients, as given by equation (80). The results for this test are presented in figure 10.

In comparison to figure 5 it is clear, that the informative prior has a positive influence on the correlation length, which in this test are maximally on the order of hundred Gibbs iterations.

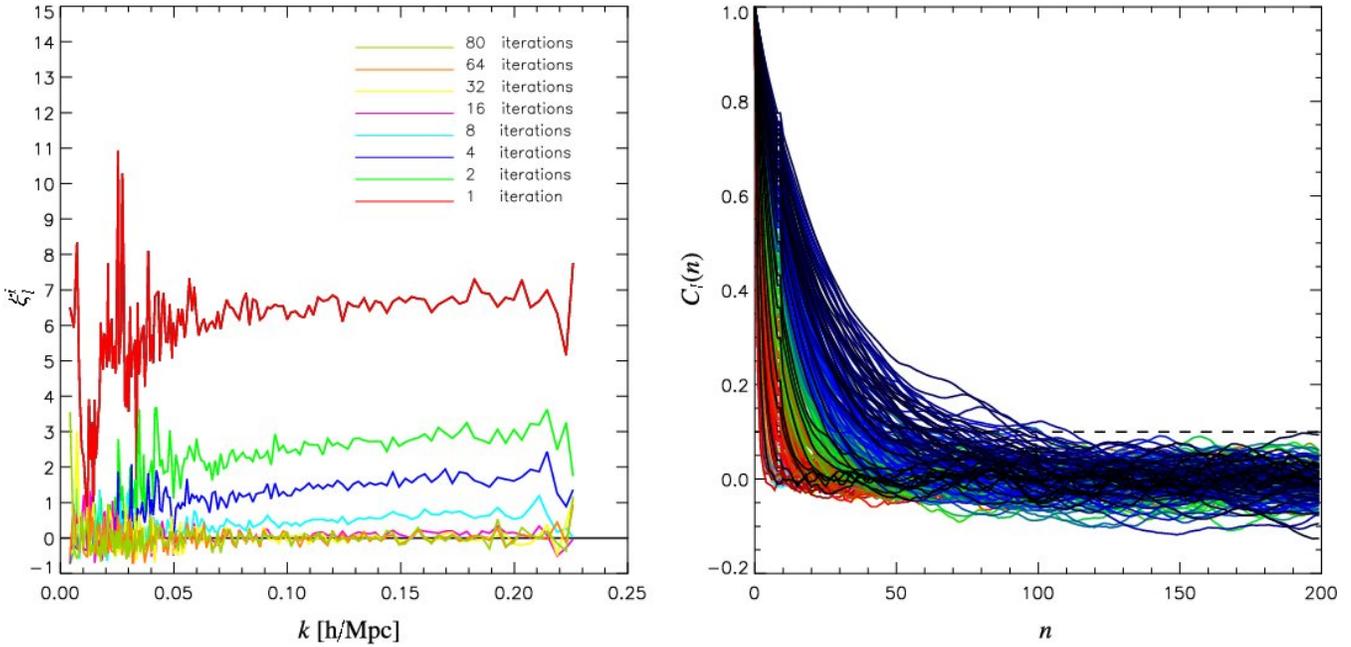
As discussed previously the long correlation length at the highest frequencies are mainly of technical nature, as the Nyquist frequencies cannot be properly represented in the finite Fourier box required for the FFT. This fact introduced artificial variance, which, however, our method can take into account. The informative prior helps in this situation, as it stabilizes these artificial fluctuations by prior information.

In addition, we observed a much better numerical behavior of our method when employing the informative prior, as the code does not run as frequently into numerically extreme regimes as with the maximum ignorance prior. This leads to a faster convergence of the conjugate gradient algorithm towards the desired accuracy.

Further, we also observe a better correction of survey geometry effects. The correlation function for the  $P_l$  is plotted in figure 11. The maximal correlation between different  $P_l$  in this test was less than 10%, which is a clear improvement.

For comparison we also plot the 2D marginalized posterior densities, for the maximally correlated and anticorrelated modes in the maximum ignorance case.

Finally, we present the low and high resolution power-spectra for the informative prior in figure 12. Note, that our prior did not contain any information on the baryonic oscillations. As can



**Figure 10.** Same as figure 5 but for low resolution runs with the inverse gamma prior.

be clearly seen, the baryonic features have nicely been recovered, demonstrating, that our informative prior provided much less information than contained in the data.

### 8.5 Testing with galaxy mock catalogs

In this section, we describe the application of ARES to a mock galaxy survey based on the Millennium run (De Lucia & Blaizot 2007).

The intention of this exercise is two-fold. First we want to test ARES in a more realistic setup, where the intrinsic shot noise of the galaxy distribution is correlated with the underlying signal, which could not be tested with the Gaussian tests before. And second, we want to demonstrate that ARES is able to reconstruct the fully evolved non-linear matter distribution of the N-body simulation.

This mock galaxy survey consists of a set of comoving galaxy coordinates distributed in a 500 Mpc box. To obtain a realistic sky observation from this full sky galaxy sample, we virtually observe these galaxies through the sky mask and according to the selection function presented in figure 3. The discrete galaxy distribution resulting from this mock observation is then sampled to a  $128^3$  equidistant grid.

To reduce gridding artifacts, such as aliasing power, we employ a supersampling technique as proposed in Jasche et al. (2009). This allows us to accurately treat the mode coupling, and will yield a precision estimate of the power-spectra up to the highest frequencies contained in the box.

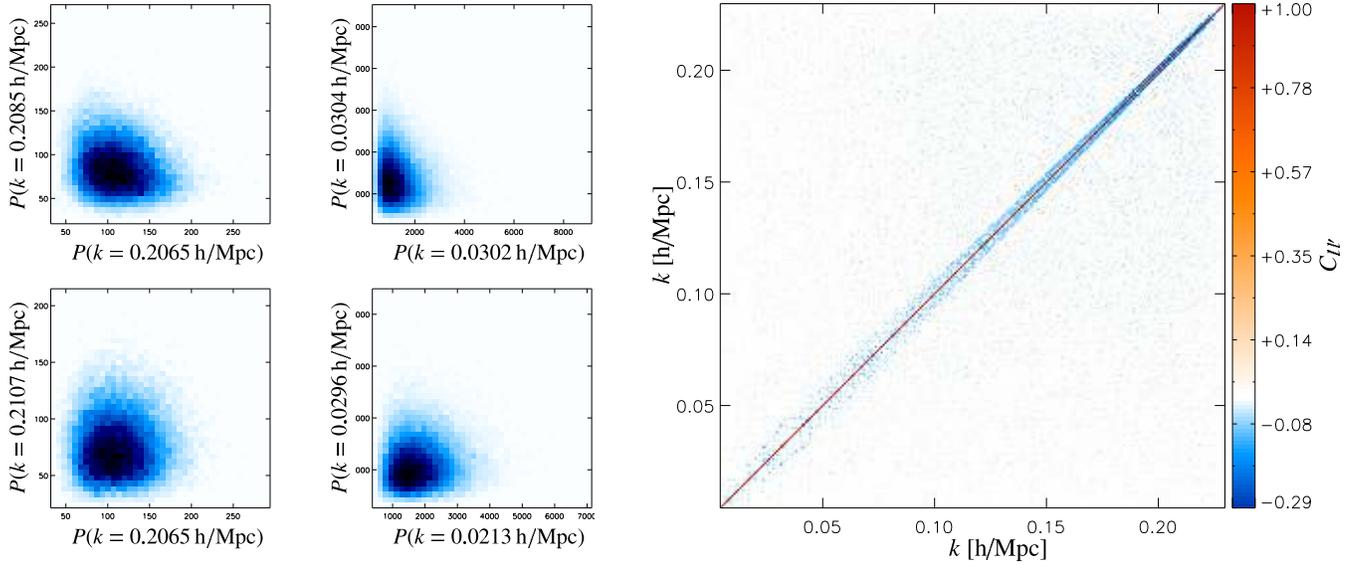
Similarly to the method described in 8.3, here we will run 4 independently initialized chains. Further, we employ the maximum ignorance Jeffrey’s prior. The galaxy distribution of this mock galaxy catalog follows the fully evolved non-linear matter distribution. Nevertheless, we initialize the Gibbs sampler with the linear power-spectrum. Then the initial burn-in period, of about 50 samples, is required to reach the non-linear power-spectrum. The systematic drift towards the correct power-spectrum is represented in figure 13. This experiment nicely demonstrates that the Gibbs sam-

pling approach is able to recover the non-linearities of the fully evolved matter density field. At this point it is important to note, that the Wiener filter is a linear operation on the data, and as such leaves the statistics of the data intact. This has been demonstrated by Kitaura et al. (2009), where they show, that the statistics of the reconstructed density field is consistent with a log-normal distribution, as expected for a non-linearly evolved matter distribution. This discussion clarifies, that the Wiener filter, or the Gibbs sampling approach as presented in this work, is very well able to capture the non-Gaussian characteristics of the density field.

However, in case one would like to perform a higher resolution analysis, it would be advantageous to initialize the Gibbs chain with a nonlinear power-spectrum guess, to yield even shorter burn-in times.

The ensemble averaged power-spectrum obtained from this run, together with the one and two sigma confidence regions, is also presented in figure 13. Here it can clearly be seen, that the recovered power-spectrum is consistent with the fully non-linearly evolved matter field. Towards the larger scales the uncertainty increase, which is due to the imposed survey geometry.

So far, in all test, we have focussed only on the recovery of the power-spectrum, and ignored the sample of reconstructed density fields. Since the Gibbs sampler also provides samples from the matter density field posterior, we are able to calculate any desired statistical summary for the matter field reconstructions. The ability to provide uncertainty estimates for the recovered density fields will in general be valuable for further science based on the matter field estimates. For this reason, in figure 14, we present the estimated mean and variance maps obtained from the 4000 Gibbs samples. As can be seen, the variance map clearly captures the features of the survey geometry and selection effects. With the set of Gibbs samples being available, all joint uncertainties can easily be propagated to the finally estimated quantities, such as the gravitational potential or large scale cosmic flows, by applying the according operation to the individual matter field samples. The result of such a procedure then again yields a probability distribution in the final



**Figure 11.** The left panel shows four 2D marginalized posterior densities, which correspond to the maximally correlated and anti-correlated cases of the previous run. The right panel shows the power-spectrum covariance matrix estimated from the 40,000 Gibbs samples calculated with the informative prior.

quantity, enabling us to provide uncertainty information for these quantities.

## 9 OPERATIONS ON THE SET OF GIBBS SAMPLES

In this section, we present an example application operating on the set of Gibbs samples.

The outcome of our Bayesian method is not a single estimate but a Gibbs sample representation of the full posterior probability distribution for the power-spectrum coefficients. We are therefore able to propagate all uncertainties to any final result, simply by applying a post processing step to all Gibbs samples. The result of such an operation would again yield a probability distribution of the estimated final quantity.

As a simple demonstration, we apply a running median filter to the set of power-spectra, which will reduce the spectral resolution.

It is known, that the median is a better estimator of the typical value of a sample than the mean when there are large extraneous outliers in the sample (Stuart & Ord 1994). For this reason, we choose the median to estimate the mode power in a given frequency bin  $\Delta k_m$ . Such a bin can be chosen to be large enough to smooth out any fluctuation below a certain scale. In our specific case we vary the bin width  $\Delta k_m$  with frequency, to allow for a logarithmic binning.

The median  $P_m^v$  of a set of power-spectrum amplitudes  $\{P_m\}$  contained within the frequency bin of width  $\Delta k_m$  around the mode  $k_m$  then satisfies the inequalities

$$\mathcal{P}(P_m \leq P_m^v) \geq \frac{1}{2} \quad (81)$$

and

$$\mathcal{P}(P_m \geq P_m^v) \geq \frac{1}{2}. \quad (82)$$

The running median is then evaluated for every frequency in the power-spectrum sample.

We apply the running median to the set of power-spectrum

samples obtained from the two Gaussian mock cases with the Jeffrey's and the inverse gamma prior. In doing so, we are able to calculate the mean and according uncertainty regions for the running median estimates. This effect has already been discussed in section 6.3. Since the reduction of frequency resolution also decreases the amount of free parameters, the total variance decreases as well.

The results of this experiment are demonstrated in figure 15. As one can easily see, the running median estimates are much smoother than the according Gibbs estimates. Also the reduction of frequency resolution by the running median estimator yields smaller confidence regions.

Finally we are interested in the recovery of the baryonic features in the power-spectrum. We therefore employ the common procedure of dividing the measured power-spectrum by one without baryonic wiggles  $P_m^{\text{nowiggles}}$ . We then obtain the wiggle function as:

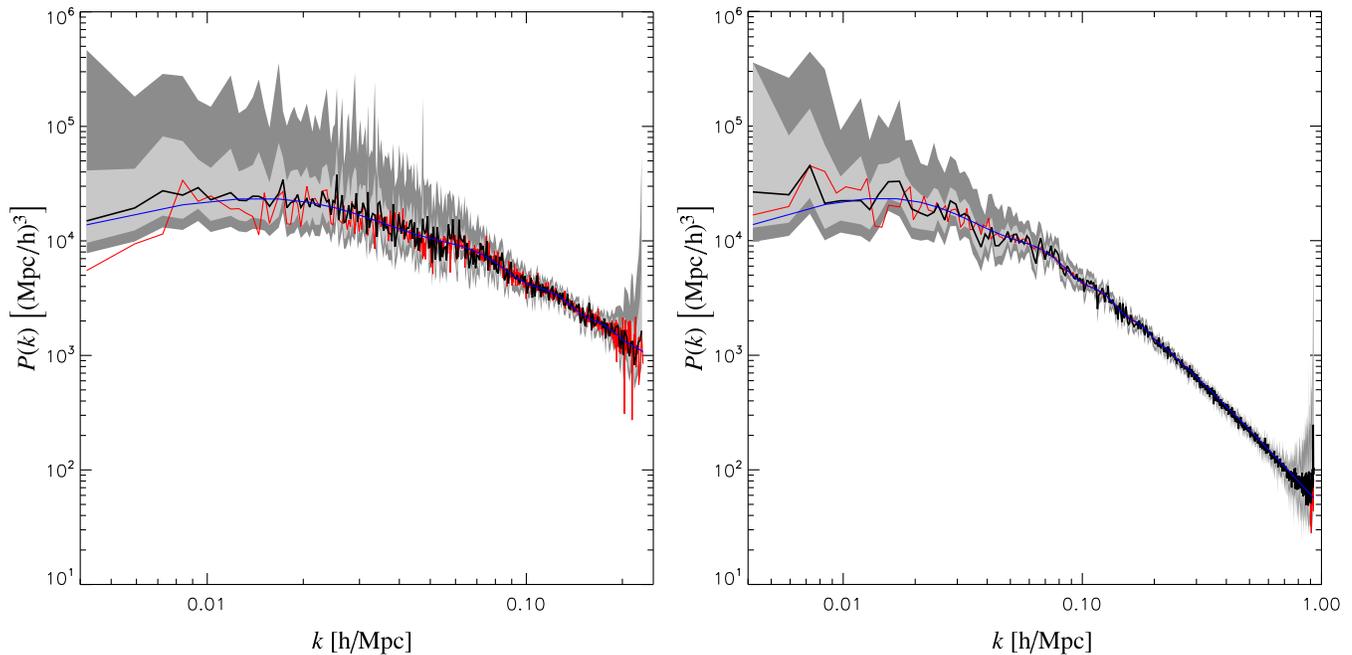
$$f_m^{\text{wiggle}} = \frac{P_m}{P_m^{\text{nowiggles}}}. \quad (83)$$

Calculating the wiggle function for all Gibbs samples and applying the running median estimator to the set of wiggle functions will yield the distribution of wiggle functions. We then estimate the mean and the according one and two sigma confidence regions. The result of this calculation is presented in figure 15 for the two Gaussian test cases. As expected, the variance towards the largest scales increases. Nevertheless, figure 15 clearly demonstrates that the baryonic features have been recovered precisely by the Gibbs sampling approach.

This example nicely demonstrates that the uncertainty estimation can easily be transported to any final quantity estimated from the set of Gibbs samples.

## 10 CONCLUSION

In this work, we presented ARES, a new Bayesian computer algorithm, designed to extract the full information on the two point statistics from any given probe of the three dimensional large scale structure.



**Figure 12.** Same as figure ?? but for the inverse gamma informative prior.

The scientific aim of this algorithm is to provide an estimate of the power-spectrum posterior  $\mathcal{P}(\{P(k_i)\}|\{d_i\})$ , conditional on a data set, while accounting and correcting for all possible sources of uncertainties, such as survey geometry, selection effects and biases. This is achieved by exploring the power-spectrum posterior  $\mathcal{P}(\{P(k_i)\}|\{d_i\})$  via a Gibbs sampling approach.

While direct sampling from the power-spectrum posterior is not possible, it is possible to draw samples from the full joint posterior distribution  $\mathcal{P}(\{P(k_i)\}, \{s_i\}|\{d_i\})$  of the power-spectrum coefficients  $P(k_i)$  and the three dimensional matter density contrast amplitudes  $s_i$  conditional on a given set of data points  $\{d_i\}$ .

The entire Gibbs sampling algorithm therefore consists of two basic sampling steps, in which first a full three dimensional Wiener reconstruction algorithm is applied to the data and then a power-spectrum is drawn from the inverse gamma distribution. In this fashion we obtain a set of power-spectrum and signal amplitude samples, which provide a full representation of the full joint posterior distribution  $\mathcal{P}(\{P(k_i)\}, \{s_i\}|\{d_i\})$ . The scientific output of this Bayesian method therefore is not a single estimate but a complete probability distribution, enabling us to calculate any desired statistical summary such as the mean, mode or variance.

We also demonstrated, that given a set of Gibbs samples, it is possible to provide an analytic approximation to the power-spectrum posterior  $\mathcal{P}(\{P(k_i)\}|\{d_i\})$ . This Blackwell-Rao estimator has an analytically appealing form enabling us to calculate any desired moment of the probability distribution in a simple analytic way.

In addition, since the full joint probability distribution is available, it is easy to carefully propagate all uncertainties through to the result of further post-processing analysis steps, such as parameter estimation.

In this work, we focused on thoroughly testing the performance and behavior of our method by applying it to simulated data with controlled properties. These tests were designed to highlight the problematic of survey geometry and selection effects, for the

two cases of Gaussian random fields and a mock galaxy catalog based on the Millennium run.

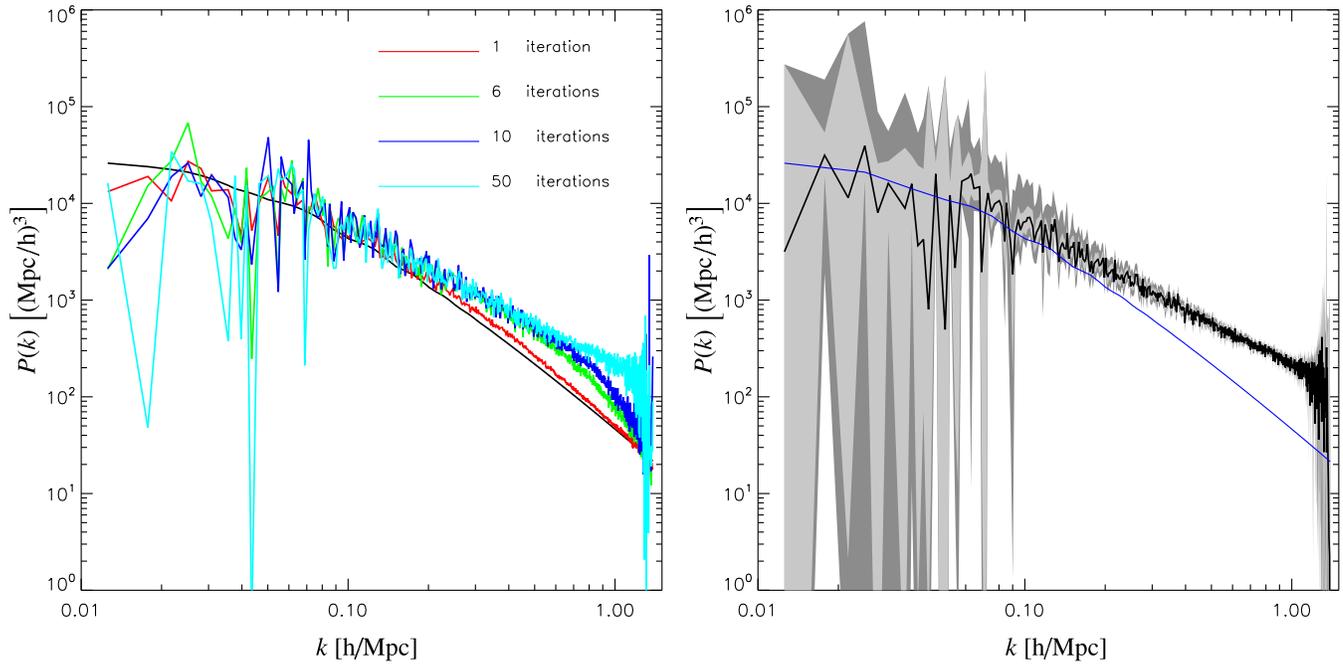
One of the main goals of these tests was to build up intuition on the phenomenological behavior of the Gibbs sampling algorithm, estimating particularly issues, such as the correlation length of the Gibbs chain, burn-in and convergence times. The result of these tests is of special relevance, as it shows how long the Gibbs sampling chain has to run in order to produce a sufficient amount of independent samples.

Through these experiments we found that the longest correlation lengths are dominated by the poorly constrained Nyquist modes of the box, which can be easily alleviated by imposing some prior knowledge on these modes. In doing so we found that the maximal correlation length for the Gibbs chain was on the order of hundred Gibbs samples. Thus, creating a large number of independent samples in a full scale data analysis is numerically very well feasible.

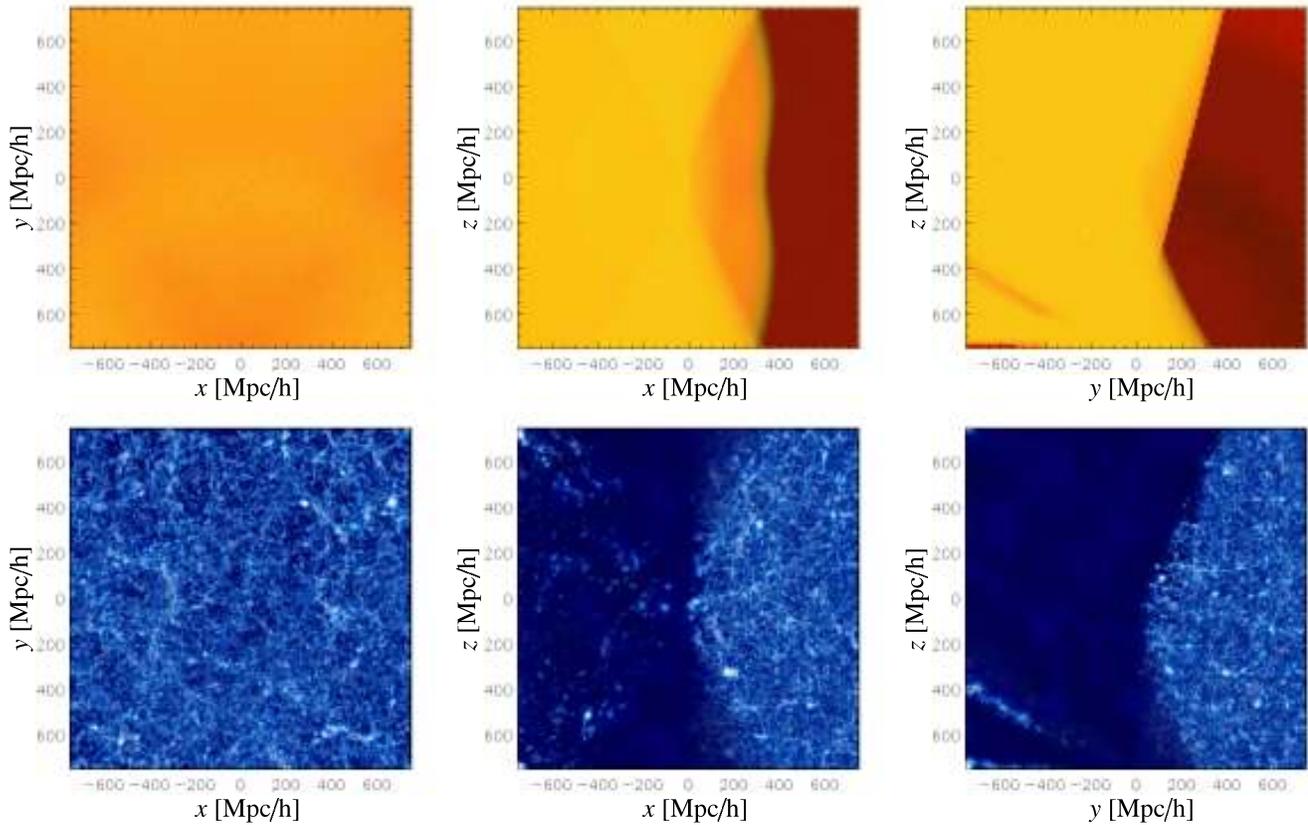
However, the most important result of these tests is, that our method is able to correct for artificial mode coupling due to the survey geometry and selection effects. This was tested by examining the correlation structure of the Gibbs samples, which showed that the maximal residual correlation can be reduced to less than 10%, demonstrating that this method correctly accounts for geometry effects.

The application of ARES to a galaxy mock catalog, based on the Millennium run, demonstrated the ability of our method to capture the characteristics of the fully nonlinearly evolved matter field. This is owed to the fact, that the Wiener filter is a linear operation on the data, and as such does not destroy the intrinsic statistical characteristics of the data set.

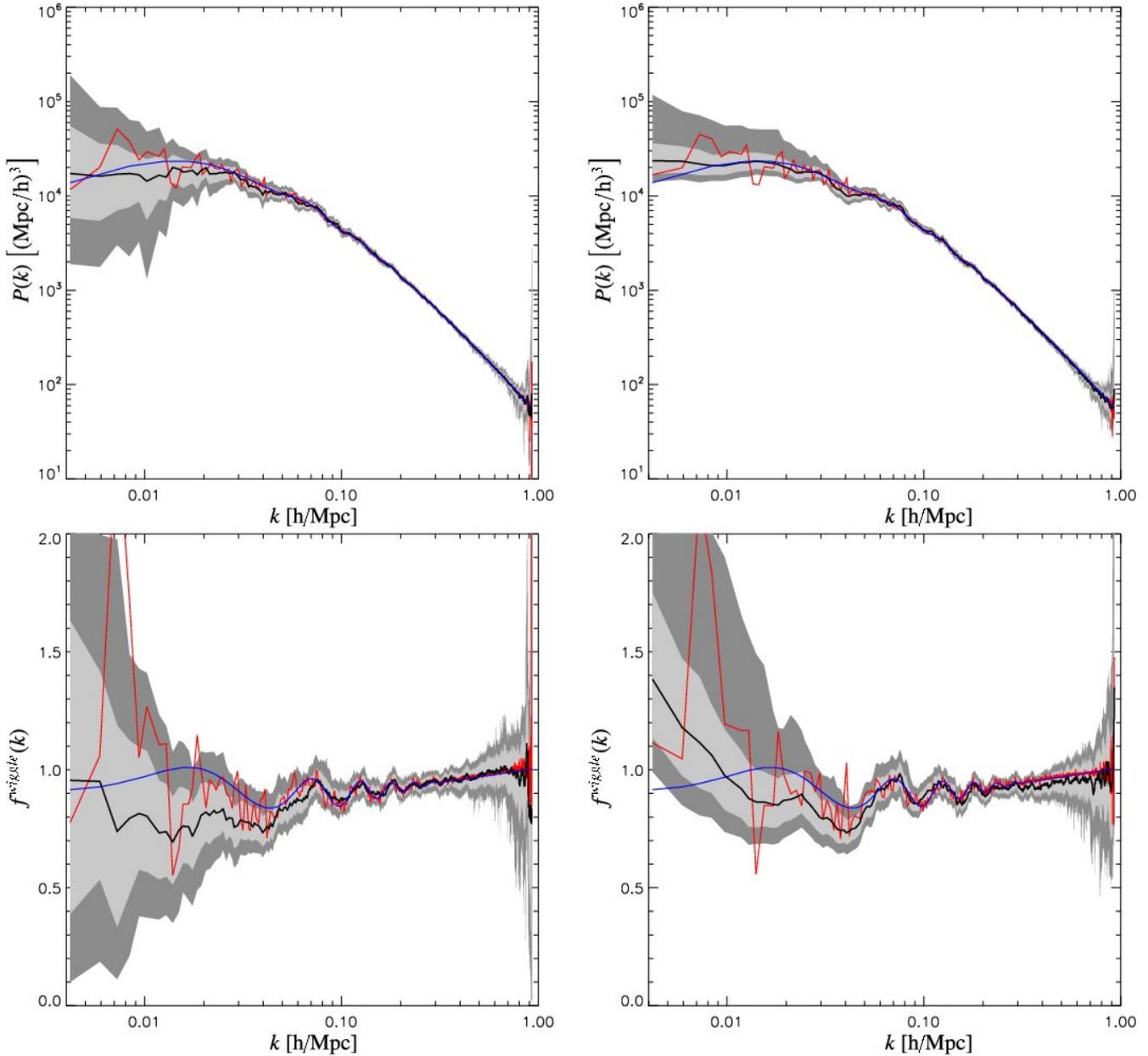
Nevertheless, a full real data analysis of existing redshift surveys requires the treatment of additional systematic effects such as scale or luminosity dependent biases or redshift space distortion corrections, which we defer to future works. However, the Bayesian framework, as presented here, can take all these effects naturally into account and treats them statistically fully consistent. Beside



**Figure 13.** The left panel shows successive power-spectrum samples during the burn-in period together with the initial guess (black line). The right panel shows the estimated power-spectrum of the Gibbs sampling analysis. The blue line denotes the initial power-spectrum guess, the black curve is the ensemble mean power-spectrum and the light gray and dark gray shaded regions represent the one sigma two sigma confidence regions respectively.



**Figure 14.** Volume rendering of the ensemble variance (upper panels) and the ensemble mean (lower panels) obtained from the mock galaxy catalog analysis.



**Figure 15.** Running median estimates of the power-spectra (upper panels) and the according wiggle functions (lower panels) for the set of Gibbs samples with the Jeffrey’s prior (left panels), and the inverse gamma prior (right panels). The black lines represent the ensemble mean of the sample set and the light gray and dark gray shaded regions denote the one and two sigma confidence regions respectively. Additionally we show the according input power-spectra. The blue line shows the cosmological power-spectrum from which the matter field realization was drawn, and the red line is the power-spectrum of this specific matter field realization.

the possibility to include various kinds of uncertainties, the Gibbs sampling approach also allows for a natural joint analysis of different data sets, taking into account the systematics of each individual data set.

In summary, we showed that ARES is a highly flexible and adaptive machinery for large scale structure analysis, which is able to account for a large variety of systematic effects and uncertainties. For this reason, ARES has the potential to contribute to the era of precision cosmology.

## ACKNOWLEDGMENTS

We are grateful to Rainer Moll and Björn Malte Schäfer for useful discussions and support with many valuable numerical gadgets. We also thank Jérémy Blaizot for several enlightening explanations and hints on the nature of galaxy distributions and the systematics of redshift surveys. Further we thank the “Transregional Collaborative Research Centre TRR 33 - The Dark Universe” and the Marie Curie FP7 fellowship for the support of this work.

## REFERENCES

- Abazajian K., et al., 2008, ArXiv e-prints
- Ballinger W. E., Heavens A. F., Taylor A. N., 1995, MNRAS, 276, L59
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
- Bertschinger E., 1987, ApJL, 323, L103
- Blake C., Glazebrook K., 2003, ApJ, 594, 665
- Cole S., et al., 2005, MNRAS, 362, 505
- Coles P., Jones B., 1991, MNRAS, 248, 1
- Colless M., et al., 2001, MNRAS, 328, 1039
- Colombi S., 1994, ApJ, 435, 536
- Cui W., Liu L., Yang X., Wang Y., Feng L., Springel V., 2008, ArXiv e-prints, 804
- De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
- Dekel A., Lahav O., 1999, ApJ, 520, 24
- Efstathiou G., et al., 2002, MNRAS, 330, L29
- Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
- Eisenstein D. J., Hu W., 1999, ApJ, 511, 5
- Eriksen H. K., O'Dwyer I. J., Jewell J. B., Wandelt B. D., Larson D. L., Górski K. M., Levin S., Banday A. J., Lilje P. B., 2004, ApJS, 155, 227
- Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
- Fisher K. B., Scharf C. A., Lahav O., 1994, MNRAS, 266, 219
- Frigo M., Johnson S. G., 2005, Proceedings of the IEEE, 93, 216
- Galassi M., Davies J., Theiler J., Gough B., Jungman G., Booth M., Rossi F., 2003, Gnu Scientific Library: Reference Manual. Network Theory Ltd.
- Ganon G., Hoffman Y., 1993, ApJL, 415, L5
- Gelfand A. E., Smith A. F. M., 1990, 85, 398
- Hamilton A. J. S., 1997a, MNRAS, 289, 285
- Hamilton A. J. S., 1997b, MNRAS, 289, 295
- Hannestad S., 2003, Journal of Cosmology and Astro-Particle Physics, 5, 4
- Heavens A. F., Taylor A. N., 1995, MNRAS, 275, 483
- Hoffman Y., Ribak E., 1991, ApJL, 380, L5
- Hu W., Eisenstein D., Tegmark M., 1998, Phys. Rev. Lett., 80, 5255
- Jasche J., Kitaura F. S., Ensslin T. A., 2009, ArXiv e-prints
- Jewell J., Levin S., Anderson C. H., 2004, ApJ, 609, 1
- Kayo I., Taruya A., Suto Y., 2001, Astrophys. J., 561, 22
- Kitaura F. S., Enßlin T. A., 2008, MNRAS, 389, 497
- Kitaura F. S., et al., 2009, Cosmic Cartography of the Large-Scale Structure with Sloan Digital Sky Survey Data Release 6, submitted to MNRAS
- Lahav O., Suto Y., 2004, Living Reviews in Relativity, 7, 8
- Martel H., 2005, ArXiv Astrophysics e-prints
- Martínez V. J., Saar E., 2002, Statistics of the Galaxy Distribution. Chapman & amp
- Matsumoto M., Nishimura T., 1998, ACM Trans. Model. Comput. Simul., 8, 3
- Michel R., Kirchoff K., 1999, Nachweis-, Erkennungs- und Vertrauensgrenzen bei Kernstrahlungsmessungen. TÜV Media GmbH TÜV Rheinland Group
- O'Hagan A., 2000, 2B
- Padmanabhan T., 1993, Structure Formation in the Universe
- Peacock J. A., Dodds S. J., 1994, MNRAS, 267, 1020
- Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815
- Percival W. J., 2005, MNRAS, 356, 1168
- Percival W. J., et al., 2001, MNRAS, 327, 1297
- Percival W. J., et al., 2002, MNRAS, 337, 1068
- Percival W. J., Verde L., Peacock J. A., 2004, MNRAS, 347, 645
- Pope A. C., et al., 2004, ApJ, 607, 655
- Sánchez A. G., Cole S., 2008, MNRAS, 385, 830
- Seljak U., 1998, ApJ, 503, 492
- Seo H.-J., Eisenstein D. J., 2003, ApJ, 598, 720
- Silk J., 1968, ApJ, 151, 459
- Spergel D., Verde L., Peiris H., Komatsu E., Nolta M., Bennett C., Halpern M., Hinshaw G., Jarosik N., Kogut A., Limon M., Meyer S., Page L., Tucker G., Weiland J., Wollack E., Wright E., 2003, Astrophys. J. Suppl. Ser., 148, 175
- Stuart A., Ord J. K., 1994, Kendall's advanced theory of statistics. Vol.1: Distribution theory. Kendall's advanced theory of statistics. Vol.1: Distribution theory, 6th ed. 3 Volumes. by A. Stuart and J.K. Ord. London: Hodder Arnold, 1994
- Sunyaev R. A., Zeldovich Y. B., 1970, A&SS, 7, 3
- Tadros H., et al., 1999, MNRAS, 305, 527
- Tanner M. A., 1996, Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, 3rd edn. Springer Series in Statistics, Springer Verlag
- Tegmark M., 1995, ApJ, 455, 429
- Tegmark M., et al. 2004, ApJ, 606, 702
- Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
- Verde L., et al., 2003, ApJS, 148, 195
- Wandelt B. D., 2004, ArXiv Astrophysics e-prints
- Yamamoto K., 2003, ApJ, 595, 577

**APPENDIX A: DISCRETE FOURIER TRANSFORMATION**

we define the numerical Fourier transformation as:

$$y_k = \hat{C} \sum_{j=0}^{N-1} x_j e^{-2\pi j k \frac{\sqrt{-1}}{N}} \quad (\text{A1})$$

and the inverse Fourier transform

$$x_j = C \sum_{k=0}^{N-1} y_k e^{2\pi j k \frac{\sqrt{-1}}{N}} = C \sum_{k=-(\frac{N}{2}-1)}^{\frac{N}{2}} y_k e^{2\pi j k \frac{\sqrt{-1}}{N}}. \quad (\text{A2})$$

The normalization coefficients  $C$  and  $\hat{C}$  are chosen such that they fulfill the requirement:

$$C\hat{C} = \frac{1}{N}. \quad (\text{A3})$$

**APPENDIX B: CHANGE TO FFT REPRESENTATION**

However, for computational convenience, we would rather like to consider the set of coefficients  $\hat{S}_{kl}$ , which are the matrix coefficients of the Fourier transformed signal covariance matrix  $S$ . The signal covariance matrix coefficients  $S_{ij}$  are related to the Fourier matrix coefficients  $\hat{S}_{kl}$  via the discrete Fourier transform defined in Appendix A:

$$\begin{aligned} S_{ij} &= C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi i k \frac{\sqrt{-1}}{N}} \hat{S}_{kl} e^{-2\pi j l \frac{\sqrt{-1}}{N}} \\ &= C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi \frac{\sqrt{-1}}{N} (ik-jl)} \hat{S}_{kl}, \end{aligned} \quad (\text{B1})$$

where  $N$  is the number of Volume cells which we consider in our calculation, and  $C$  is the normalization of the discrete Fourier Transform. The matrix elements of the Jacobi matrix for this coordinate transformation can then be written as:

$$\begin{aligned} \mathcal{J}_{(ij)(mn)} &= \frac{\partial S_{ij}}{\partial \hat{S}_{mn}} = C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi \frac{\sqrt{-1}}{N} (ik-jl)} \frac{\partial \hat{S}_{kl}}{\partial \hat{S}_{mn}} \\ &= C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi \frac{\sqrt{-1}}{N} (ik-jl)} \delta_{km}^k \delta_{ln}^k \\ &= C^2 e^{2\pi \frac{\sqrt{-1}}{N} (im-jn)} \\ &= C^2 A_{(ij)(mn)}, \end{aligned} \quad (\text{B2})$$

where we defined the transformation matrix  $A_{(ij)(mn)} \equiv e^{2\pi \frac{\sqrt{-1}}{N} (im-jn)}$ . The norm of the Jacobi determinant can then be calculated in matrix notation as:

$$\begin{aligned} |\det(\mathcal{J})| &= |\det(C^2 A)| \\ &= |\det(C^2 \mathbb{I}_{N^4} A)| \\ &= |\det(C^2 \mathbb{I}_{N^4})| |\det(A)| \\ &= C^{2N^4} |\det(A)|, \end{aligned} \quad (\text{B3})$$

where we made use of the fact, that the Jacobi matrix is a quadratic  $N^2 \times N^2$  matrix, which is true for usual Fast Fourier Transform implementations. Since  $A$  is a unitary matrix:

$$(A_{(ij)(mn)})^\dagger = A_{(mn)(ij)}^*$$

$$\begin{aligned} &= e^{-2\pi \frac{\sqrt{-1}}{N} (mi-nj)} \\ &= e^{-2\pi \frac{\sqrt{-1}}{N} (im-jn)} \\ &= A_{(ij)(mn)}^{-1}, \end{aligned} \quad (\text{B4})$$

the norm of the determinant  $\det(A)$  is unity. Therefore, the norm of the Jacobi determinant is just a constant:

$$|\det(\mathcal{J})| = C^{2N^4}. \quad (\text{B5})$$

With this definition we can perform the change of coordinates and express the probability distribution given in equation (??) in terms of the set of the matrix coefficients  $\hat{S}_{kl}$ :

$$\mathcal{P}(\{\hat{S}_{kl}\}|\{s_i\}) = C^{2N^4} \mathcal{P}(\{S_{ij}\}|\{s_i\}) = C^{2N^4} \frac{\mathcal{P}(\{\hat{S}_{kl}\})}{\mathcal{P}(\{s_i\})} \mathcal{P}(\{s_i\}|\{\hat{S}_{kl}\}), \quad (\text{B6})$$

where  $\mathcal{P}(\{\hat{S}_{kl}\})$  is the signal covariance matrix prior in Fourier space, and  $\mathcal{P}(\{s_i\}|\{\hat{S}_{kl}\})$  is given as:

$$\begin{aligned} \mathcal{P}(\{s_i\}|\{\hat{S}_{kl}\}) &= \mathcal{P}(\{s_i\}|S) \\ &= \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{1}{2} \sum_i \sum_j s_i S_{ij}^{-1} s_j} \\ &= \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{1}{2} \sum_i \sum_j s_i C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi \frac{\sqrt{-1}}{N} (ik-jl)} \hat{S}_{kl}^{-1} s_j} \\ &= \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{C^2}{2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sum_i s_i e^{2\pi \frac{\sqrt{-1}}{N} ik} \hat{S}_{kl}^{-1} \sum_j s_j e^{-2\pi \frac{\sqrt{-1}}{N} jl}} \\ &= \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{C^2}{2C^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \hat{S}_k^* \hat{S}_k^{-1} \hat{s}_l}, \end{aligned} \quad (\text{B7})$$

with  $\hat{s}_l$  being the signal coefficients of the discrete Fourier transformed signal. To determine the determinant  $\det(2\pi S)$  in Fourier space, we rewrite equation (B2) as a matrix product:

$$\begin{aligned} S_{ij} &= \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} C e^{2\pi i k \frac{\sqrt{-1}}{N}} \hat{S}_{kl} C e^{-2\pi j l \frac{\sqrt{-1}}{N}} \\ &= \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} C A_{ik} \hat{S}_{kl} C A_{lj}^* \end{aligned} \quad (\text{B8})$$

where we defined the unitary matrix  $A_{ik} \equiv e^{2\pi i k \frac{\sqrt{-1}}{N}}$ . The determinant can then be evaluated in matrix notation as:

$$\begin{aligned} \det(2\pi S) &= \det(2\pi C^2 A \hat{S} A^{-1}) \\ &= \det(C^2 \mathbb{I}_{N^2} A 2\pi \hat{S} A^{-1}) \\ &= \det(C^2 \mathbb{I}_{N^2}) \det(A) \det(2\pi \hat{S}) \det(A^{-1}) \\ &= C^{2N^2} \det(2\pi \hat{S}). \end{aligned} \quad (\text{B9})$$

**APPENDIX C: WIENER VARIANCE**

In section 4.3 we described, that a signal sample can be obtained by adding a fluctuation term  $y_i$  to the Wiener mean reconstruction  $m_i$ . According to equations (36) and (40) we can rewrite the fluctuation term as:

$$\begin{aligned} y_i &= s_i^* - \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} d_l^* \\ &= s_i^* - \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} \left( \sum_k K_{lk} s_k^* + \epsilon_l^* \right) \end{aligned}$$

$$\begin{aligned}
 &= \sum_j D_{ij}^{-1} \left[ \sum_k D_{jk} s_k^* - \sum_m \sum_l K_{mj} N_{ml}^{-1} \left( \sum_k K_{lk} s_k^* + \epsilon_l^* \right) \right] \\
 &= \sum_j D_{ij}^{-1} \left[ \sum_k \left( D_{jk} - \sum_m \sum_l K_{mj} N_{ml}^{-1} K_{lk} \right) s_k^* - \sum_m \sum_l K_{mj} N_{ml}^{-1} \epsilon_l^* \right] \\
 &= \sum_j D_{ij}^{-1} \left[ \sum_k S_{jk}^{-1} s_k^* - \sum_m \sum_l K_{mj} N_{ml}^{-1} \epsilon_l^* \right]. \quad (C1)
 \end{aligned}$$

From this we immediately see that the mean of  $y_i$ :

$$\langle y_i \rangle = \sum_j D_{ij}^{-1} \left[ \sum_k S_{jk}^{-1} \langle s_k^* \rangle - \sum_m \sum_l K_{mj} N_{ml}^{-1} \langle \epsilon_l^* \rangle \right] = 0 \quad (C2)$$

vanishes by construction, as  $\langle s_k^* \rangle = \langle \epsilon_l^* \rangle = 0$ . Note, that the signal and noise covariance are given by:

$$\langle s_i^* s_j^* \rangle = S_{ij} \quad (C3)$$

and

$$\langle \epsilon_i^* \epsilon_j^* \rangle = N_{ij}. \quad (C4)$$

Also note that, by construction, there is no correlation between the two random variates  $s_i^*$  and  $\epsilon_i^*$  as they have been created by two independent random processes. Therefore, we yield:

$$\langle s_i^* \epsilon_j^* \rangle = 0. \quad (C5)$$

Then we can calculate the variance as:

$$\begin{aligned}
 \langle y_i y_q \rangle &= \left\langle \sum_j D_{ij}^{-1} \left[ \sum_k S_{jk}^{-1} s_k^* - \sum_m \sum_l K_{mj} N_{ml}^{-1} \epsilon_l^* \right] \sum_r D_{qr}^{-1} \left[ \sum_s S_{rs}^{-1} s_s^* - \sum_t \sum_u K_{tr} N_{tu}^{-1} \epsilon_u^* \right] \right\rangle \\
 &= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} \left[ \sum_{ks} S_{jk}^{-1} S_{rs}^{-1} \langle s_k^* s_s^* \rangle + \sum_{mltu} K_{mj} N_{ml}^{-1} K_{tr} N_{tu}^{-1} \langle \epsilon_l^* \epsilon_u^* \rangle \right] \\
 &= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} \left[ \sum_{ks} S_{jk}^{-1} S_{rs}^{-1} S_{ks} + \sum_{mltu} K_{mj} N_{ml}^{-1} K_{tr} N_{tu}^{-1} N_{lu} \right] \\
 &= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} \left[ \sum_k S_{jk}^{-1} \delta_{rk}^K + \sum_{mlt} K_{mj} N_{ml}^{-1} K_{tr} \delta_{tl}^K \right] \\
 &= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} \left[ S_{jr}^{-1} + \sum_{ml} K_{mj} N_{ml}^{-1} K_{lr} \right] \\
 &= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} D_{jr} \\
 &= \sum_r D_{qr}^{-1} \delta_{ir}^K \\
 &= D_{iq}^{-1}. \quad (C6)
 \end{aligned}$$

Therefore, the additive fluctuation term  $y_i$  provides the correct variance.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.