Bayesian analysis of spatially distorted cosmic signals from Poissonian data

Cornelius Weig* and Torsten A. Enßlin
Max-Planck-Institut für Astrophysik (MPA), Karl-Schwarzschildstr. 1, D-85748 Garching, Germany

Accepted 2010 May 31. Received 2010 May 3; in original form 2010 March 16

ABSTRACT
Reconstructing the matter density field from galaxy counts is a problem frequently addressed in current literature. Two main sources of error are shot noise from galaxy counts and insufficient knowledge of the correct galaxy position because of peculiar velocities and redshift measurement uncertainty. Here, we address the reconstruction problem of a Poissonian sampled lognormal density field with velocity distortions in a Bayesian way using a maximum a posteriori method. We test our algorithm on a one-dimensional test case and find significant improvement compared to simple data inversion. In particular, we address the following problems: photometric redshifts, mapping of extended sources in coded mask systems, real space reconstruction from redshift space galaxy distribution and combined analysis of data with different point spread functions.

Key words: methods: data analysis – techniques: photometric – dark matter – large-scale structure of Universe.

1 INTRODUCTION
Correct and thorough signal analysis is vital in cosmology. This is for one thing because of the often low signal-to-noise ratios of many cosmic measurements. An important fact to notice here is that many measurements cannot be independently repeated, as nature only grants us with one realization of the data, such as the cosmic microwave background (CMB) or the large-scale structure of the Universe (LSS).

Because the complications in extracting the desired information from data are so fundamental, it is often not possible to draw sensible conclusions from the data without some knowledge of the properties of the underlying signals. Although it would be desirable to be completely independent and prejudice-free, typically some freedom has to be given up in the analysis by restricting to a specific model. In return, much better constraints on the measured quantity are obtained. Hence, a thorough signal analysis must take care of all these aspects, otherwise the wrong conclusions could be drawn.

This is most naturally done in the Bayesian framework where all variables are considered to be subject to error and variance.

The emphasis in this paper is on the reconstruction problem of a lognormal density field, which is sampled using a Poissonian process as a simple description of galaxy formation and a number of other processes, such as a highly structured \( \gamma \)-ray emissivity. We choose the lognormal field, because we think it suited to model the dark matter density distribution of the Universe (see Section 3.1.1).

In addition, we extend the problem such that the signal is spatially distorted and galaxy counts from one position may show up at other locations. This way, a sharp peak in the signal can show up as a broad distribution in the data depending on the underlying distortion process. This allows us to address the real space reconstruction problem of the dark matter density field from redshift distorted galaxy counts and to naturally incorporate photometric redshift errors in our analysis, to name a few.

The most generic case of uncertain position is the measurement error from the measurement apparatus itself, which comes with any measurement. In many cases, these errors form a Gaussian distribution around a mean value. However, there are also other cases, such as photometric redshift, where because of the measurement technique there is a considerable chance of ‘catastrophic outliers’, which leads to non-Gaussian probability distributions. In our analysis, the chance of such catastrophic errors can be naturally included and dealt with. This permits us to deal with cases where such outliers are the rule, for example coded mask detectors in X-ray and \( \gamma \)-ray astronomy where a point source has to be identified via its complex mask shadow on the detector plane. Other areas where spatial distortion should be included in the analysis are \( \gamma \)-ray astronomy using Čerenkov telescopes, or ultra-high-energy cosmic ray detectors, because of the extended point spread functions (PSFs) of the measurement devices.

In all examples so far, the distortion of the data was known a priori and was fixed. However, we can go one step further and allow the distortion to depend on the signal itself that we want to measure. The paradigm for this problem is the measurement of redshift in galaxy surveys. Here, the aim is to measure the real space density
distribution of matter in the Universe using galaxy counts. However, as the presence of matter has an effect on the peculiar velocities of the observed galaxies, the distortion depends on the details of the signal to be reconstructed. Beyond the linear regime, where a dark matter halo has collapsed to a virialized object such as a galaxy cluster, the galaxies have large peculiar velocities. Because only the component along the line of sight adds to the redshift, those collapsed objects appear as dense elongated structures in redshift space pointing towards the observer. Therefore, this is called the ‘finger-of-god’ effect.

Including the feedback of matter on the redshift space distortions is the ultimate goal of LSS reconstruction. The point of our work is to address one very important effect so far often ignored in statistical inference of the LSS: spatial distortions. In order to focus our discussion on this, we work with simplified descriptions of the complex galaxy formation process. The adopted description, however, has previously been shown to provide good reconstructions despite its simplicity.

Significant progress in the field of large-scale structure reconstruction with a lognormal model for the dark matter overdensity has been achieved in a number of recent works. Enßlin, Frommert & Kitaura (2009) have derived the maximum a posteriori (MAP) estimator and loop corrections thereof within the information field theory (IFT) framework. Kitaura, Jasche & Metcalf (2010) have successfully applied the MAP Poissonian lognormal filter on mock data from N-body simulations. Jasche et al. (2010) have even achieved a reconstruction from real world Sloan Digital Sky Survey (SDSS) Data Release 7 (DR7) data going beyond the MAP approximation by using a Hamiltonian sampling method. However, none of these approaches takes the spatial uncertainty of redshift measurements (i.e. the PSF) into account. The complications from a non-trivial PSF have been addressed in a number of works with similar data models as ours. Hebert & Leahy (1992), Green (1990) and Wang, Fu & Qi (2008) have considered a similar data model but used a signal clique prior adequate for image reconstruction. Oh & Frieden (2009) had a distorted Poissonian data model but used a smoothing prior based on Fisher information for the signal. Nunez & Llacer (1990) also worked with a Poissonian data model but used an ad hoc image entropy as prior for their signal. Frieden & Wells (1978) and Gull & Daniell (1978) worked with maximum entropy prior for the signal distorted by a PSF and approximated the Poissonian distribution by a Gaussian. Different choices for such entropy priors are discussed in Nityananda & Narayan (1982) and Cornwell & Evans (1985).

The outline of this paper is as follows. We formulate the Bayesian reconstruction problem in a field-theoretic language in Section 2. In Section 3 we address the reconstruction problem from spatially distorted data. A suitable data model for this purpose is introduced. In particular, this includes a discussion of the distortion operator in Section 3.1.2.3. We address how approximate error bars can be constructed for our reconstructions. In Section 3.3 we apply our method to the LSS reconstruction from photometric redshift measurement in a one-dimensional (1D) test case. We show how data sets with different error characteristics can be naturally combined within our framework in Section 3.4. In Section 3.5, we apply the same algorithm to a completely different field, namely X-ray and γ-ray astronomy via coded mask telescopes. Finally, in Section 3.6 we formulate a distortion problem where the distortion itself depends on the signal to be reconstructed. We therefore propose a model of how to approach the forward problem to transform from real into redshift space. We compare our results for this distortion model to a Metropolis–Hastings sampling method in Section 3.6.4.

In Section 4, we summarize our findings. Details about the notation can be found in Appendix A.

2 THEORETICAL FRAMEWORK

2.1 The way from data to information

There is always a difference between the data that we measure and the information we want to extract from the data. The process from mere data acquisition to information extraction needs a model for the way data are produced from a signal s. In a practical application, we measure data d and try to infer the signal that produced the data. In most cases, this data inversion is not unambiguous when, for example, the signal s is continuous and the data d are discrete. Then, it is necessary to formulate the data model as probability distribution functions (PDFs) such as P(d|s), P(s), P(d) and P(s|d), which we call likelihood, prior, evidence and posterior, respectively, following standard Bayesian naming conventions.

This model may include both physical and technical processes, such as noise from the measurement system or natural and unavoidable sources of error. Noise in particular makes the mapping from signal to data ambiguous and non-deterministic, such that different signals can produce the same data.

Although it has been shown that it is in principle possible to extract prior information on the signal s from the data (Enßlin & Frommert 2010), we always assume that P(s) is given in advance. In many situations, the prior is taken to be flat with the intention of being ‘prejudice-free’. However, the flatness of P(s) depends on the coordinate system chosen for s and is therefore a (hidden) choice for a specific coordinate system. Hence, working with an explicit prior should not be seen as a flawed bias for a specific model but as a complete discussion of the problem.

2.2 Optimal map-making

Unfortunately, in most situations the process of data generation from the true signal st is not reversible, which means that some information is irreversibly lost and st cannot be fully restored from the data. So, the task of signal inference from a data set d is to produce a map m as an approximation of st provided the likelihood P(d|s) how data emerge from the signal and the prior P(s). A reasonable map-making strategy is to minimize the average error that is made in the reconstruction of st. Here, it is of key importance how the errors are weighted. A suitable measure for the error is the L2-norm defined as

$$
\|s\|_2^2 = \left( \int dx \, |s(x)|^2 \right)^{1/2},
$$

where the appearing integral is a volume integral over the whole position space. Minimizing the expected error \(\langle \|s - m\|_2^2 \rangle_{\Omega|d}\) provides the prescription

$$
m_s = (s^*_s)_{\Omega|d} \equiv \int Ds s P(s|d),
$$

for map-making, where the subscript s refers to the point or pixel whose estimated map value is to be calculated. As s is a field, the \(Ds\) integral is a path integral, which runs over all possible field configurations. Usually, we have to take thorough care about the convergence of path integrals, but in all practical applications we deal with a finite number of signal points or pixels, such that the path integral reduces to a volume integral over a rather high but finite dimensional space.

© 2010 The Authors. Journal compilation © 2010 RAS, MNRAS 409, 1393–1411

by guest on 28 July 2018

Downloaded from https://academic.oup.com/mnras/article-abstract/409/4/1393/984616
Of course, different $L_p$-norms lead to different map-making techniques. Thus, it is a question of taste and belief which $L_p$-norm is preferred and therefore which map-making technique is considered best. However, as we know that the reconstruction cannot fully recover the exact signal, it makes sense to be generous to small deviations but penalize large deviations strongly, which is exactly what the $L_2$-norm does.

As a word of caution, it should be mentioned here that although equation (2) is independent of the coordinate system chosen for the data $d$, it does however depend on the coordinates of $s$. Thus, we need to specify in which signal coordinates we want to minimize the reconstruction error.

2.3 Information field theory formulation of moment calculation

Moment calculation can also be formulated in the language of statistical field theory and as we are dealing with information here, we call this framework IFT. This has been addressed, for example, by Enßlin et al. (2009), Bialek, Callan & Strong (1996), Lemm (1999, 2001) and Lemm, Uhlig & Weiguny (2001, 2005), but for completeness we briefly summarize their findings. Bayesian law allows us to rewrite equation (2) as

$$\langle s_x \rangle_{i(d)} = \int Ds \frac{P(s) P(d|s)}{P(d)} = \int Ds \frac{P(s, d)}{P(d)}.$$  

As the set of all possible signals $s$ is exclusive and exhaustive, we can marginalize $P(s, d)$ over $s$

$$P(d) = \int Ds P(s, d).$$

So it turns out that the only required probability distribution for moment calculation is $P(s, d)$.

The crucial step now is to realize that the above integrals can also be formulated in the language of statistical field theory as, for example, Lemm (1999, Enßlin et al. 2009), Jaynes (1957) and others proposed. Following, amongst others, the idea of Metropolis et al. (1953), Hastings (1970), Geman & Geman (1984) and Duane et al. (1987), we define a probability Hamiltonian as

$$H_d[s] = -\log P(s, d),$$

which leads to the IFT equivalent of the partition function

$$Z_d = \int Ds e^{-H_d[s]}.$$  

Comparing equation (6) with equation (4) we recognize that $P(d) = Z_d$. So the posterior can be expressed as $P(s|d) = e^{-H_d[s]} / Z_d$.

The full advantage of this formulation becomes evident when the posterior is or is approximated by a Gaussian, i.e.

$$P(s|d) = \mathcal{G}(s - \mu, D) \equiv \frac{1}{\sqrt{2\pi}D} e^{-(s-\mu)^2 / 2D}.$$  

Then, we can analytically calculate the generating functional

$$Z_d[J] = \int Ds e^{-H_d[s] - J^s} = e^{\mu^s J^s + s^s J^s / 2},$$

where $J$ is an arbitrary field that we call the source field in analogy to quantum field theory. From $Z_d[J]$, all moments can be calculated by functional derivation:

$$\langle s_{x_1} \cdot \cdot \cdot s_{x_n} \rangle_{i(d)} = \left. \frac{\delta^n Z_d[J]}{\delta J_{x_1} \cdot \cdot \cdot \delta J_{x_n}} \right|_{J = 0}.$$  

This result is useful in Section 3.2.3 where we address error estimation.

2.4 Maximum a posteriori approximation

For many applications, calculating the partition function is not feasible, because the joint probability is too complex for the generating functional to be calculated analytically. In these cases, we have to resort to an approximation of $\langle s_{x_1} \rangle_{i(d)}$, such as the MAP map $m_{cl}$, which approximates the posterior mean by the map that maximizes $P(s|d)$. The evidence only serves as a normalization constant, so maximizing the posterior comes down to maximizing the joint probability of signal and data. As $P(s, d)$ can be expressed in terms of $H_d[s]$, maximizing $P(s, d) = e^{-H_d[s]}$ is (because of the monotony of the exponential function) equivalent to minimizing the Hamiltonian. Minimizing the Hamiltonian is a well-known principle from classical mechanics, so it makes sense to call the MAP the classical map.

The posterior average $\langle s \rangle_{i(d)}$ and $m_{cl}$ are exactly equal, if $P(s - m_{cl}|d)$ is a symmetric and singly peaked function. In the more interesting cases, however, this is usually not the case. Yet $\langle s \rangle_{i(d)} \approx m_{cl}$ often holds when $P(s|d)$ is sharply peaked. While the posterior mean minimizes the expected $L_2$-distance from reconstruction to the true signal, we can show that the classical map minimizes the $L_1$-distance.

3 RECONSTRUCTION OF SPATIALLY DISTORTED SIGNALS

3.1 Signals with uncertain position measurement

The reconstruction of a signal from data with distorted or imprecise position information is a generic problem class.

By ‘signal’ we mean a distributed physical quantity (i.e. a field). In principle, the signal must be considered to be continuous, but for most practical applications it must be sampled, which yields a field vector $s$. Therefore, we assume from now on that $s$ is indeed an ordinary vector from a high-dimensional real vector space. We also call this space of all possible signals the signal phase space, its elements $s$ a signal realization and the value of $s$ at location $i$ the field strength $s_i$.

The signal can be probed by a measurement apparatus, which ultimately produces data that are discrete by nature. From this alone it is clear that data and signal are a priori defined on different vector spaces. In the following, the data are always the counts rates of photons in an X-ray detector. The number of such events as a function of photons in an X-ray detector. The number of such events as a vector.

The response $\mu$ of an experimental set-up is by definition the expectation value of the data $d$ averaged over all possible data realizations:

$$\mu \equiv \langle d \rangle_{i(d)}. $$  

In other words, the response is perfect noiseless data. We say that the response is local, when the field strength $s_i$ triggers only events in one data bin $d_i$. If, however, $s_i$ may trigger events in a number of data bins $d_i, d_{i+1}, \ldots$ we say that the response is non-local. If, in addition, the same data bin $d_i$ may be filled by different $s_i, s_0$ we call the data space distorted. We address both the problems, where the distortion is independent of the underlying signal and also where the distortion does depend on the signal.

In this paper, we always assume that the signal has a lognormal PDF with known covariance. We choose this signal, as we believe
that it approximately models the large-scale matter distribution and other signals.

3.1.1 Lognormal distribution for matter

There are several good reasons to believe that the lognormal PDF models the large-scale matter distribution well. For one thing, it has already been used by Hubble (1934) as early as 1934, to successfully model the galaxy count rates in two-dimensional sky patches. For another, Coles & Jones (1991) found that if an initially Gaussian random field, as is predicted by most inflationary models and more importantly as is observed in the CMB, evolves over time and the peculiar velocities grow linearly, then the initial Gaussian field is evolved over time into a lognormal field.

There is also evidence from N-body simulations that the lognormal PDF is an adequate description of the large-scale matter distribution. By direct comparison of the one-point and two-point correlation functions obtained from N-body simulations to the one-point and two-point lognormal PDFs, Kayo, Taruya & Suto (2001) found that the former can be very accurately modelled by the latter even in the strongly non-linear regime with overdensities up to 100. Kitaura et al. (2010) could even show with mock tests that their Poissonian lognormal model was able to reconstruct the matter distribution accurately for overdensities up to 1000. Furthermore, Neyrinck, Szapudi & Szalay (2009) have also shown with data from N-body simulations that a lognormal density field fits the density power spectrum much better in the slightly non-linear regime than Gaussian fields do. In both of the above works, the N-body simulations were seeded with small Gaussian density fluctuations. Recently, the lognormal model has also been successfully applied to matter density reconstruction from SDSS data (Jasche et al. 2010; Jasche & Kitaura 2010).

Last but not least, the lognormal PDF is mathematically simple and therefore comparatively easy to handle. Hence, we assume that the matter density \( \rho \) in the Universe can be modelled by

\[
\rho \equiv \rho_0 e^s. \tag{11}
\]

Here, \( \rho_0 \) is a reference density, close to but different from the mean density, the exponential function is meant to be taken component-wise and the signal \( s \) is a field with Gaussian covariance matrix \( S \) [i.e. \( P(s) = \mathcal{N}(s, S) \)]. Throughout this paper, we always assume that \( s \) is a Gaussian field with covariance matrix \( S \), if not stated otherwise.

One characteristic of the lognormal distribution is that it generates very large overdensities in the case of large \( S \), while the inner structure of void regions is barely visible to the eye. Besides, the log-density field contains information about the primordial density fluctuations. Therefore, instead of reconstructing the density field itself, we reconstruct the log-density field \( s = \log(\rho/\rho_0) \), which is Gaussian for the lognormal distribution.

3.1.2 Data model

3.1.2.1 Local response. As argued in Section 2.1, the data model must provide a likelihood to obtain some data \( d \) given some signal \( s \). As only a minor portion of matter radiates, we have to rely on tracers of matter density. It is widely accepted that galaxy count rates can serve as tracers for the dark matter density field on large scales.

As a first step, we have to find a formulation for the local response to the signal. Therefore, we subdivide the observed space into boxes of volume \( V_i \) and relate \( \mu_i \), the expected number of events within \( V_i \), to the field strength \( s_i \), the signal strength averaged over \( V_i \). We assume the local response \( \mu \) of the observation to be given by

\[
\mu_i[s] = w_i V_i \rho_i^{(0)} \left( \frac{\rho_i}{\rho_0} \right)^b = k_i e^{bs_i}. \tag{12}
\]

Here, \( w \) is the window function, which encodes information on the exposure time and survey geometry that might have an impact on the detection probability in \( V_i \), and \( \rho_i^{(0)} \) is some reference event density. For brevity, we have defined \( k_i = w_i V_i \rho_i^{(0)} \), which we call zero-response, because \( \mu(s = 0) = k \). Throughout this paper, we always assume that \( k \) is known a priori.

The bias \( b \) determines how fast and how strong overdensities produce events. There are strong reasons to assume that a single linear bias factor is an oversimplification of nature as, for example, Fry & Gaztanaga (1993), Matsubara (2008a,b) and Jeong & Komatsu (2009) have shown. Nevertheless, for the proof of concept at which we aim, the single bias factor simplifies the set-up and preserves the relevant features of more elaborate models. Besides, a scale-dependent bias for \( s \) can be incorporated in the covariance matrix by working in the Fourier picture (Kitaura et al. 2010). For \( b \ll 1 \), the response can be expanded as a power series in \( s \)

\[
\mu_i[s] = k_i \sum_{j=0}^{\infty} \frac{(bs_i)^j}{j!} \approx k_i(1 + bs_i),
\]

which is the familiar form of a linear bias model used in galaxy cosmology. However, when \( b \) approaches unity for a signal with \( O(1) \) variance, higher orders of \( s \) become more and more important and the linear approximation is bound to fail.

It must be stressed that \( \rho_i^{(0)} \) is not the average event density, because

\[
\bar{\rho}_{ei} \equiv \langle \mu_i \rangle_{(0)} = k_i \int \mathcal{D}x \ e^{bs_i} \mathcal{G}(s, S) = k_i e^{bs_0/2}
\]

(e.g. Coles & Jones 1991; Kayo et al. 2001). When setting up a simulation, we usually specify \( \bar{\rho} \), and not \( \rho_i^{(0)} \), so we have to keep this relation in mind when the bias is varied.

3.1.2.2 Poissonian sampling process. So far, we have only defined the response, which specifies how many events we can expect on average in \( V_i \) provided the signal \( s \). However, as the number of events in \( V_i \) is always a (non-negative) natural number, we obtain the true response only if the events are averaged over many different data realizations. However, in many cases nature prohibits this practice, because the number of events \( d_i \) is a random number with expectation value \( \mu_i \) that is not redrawn between observations, such as the random number of galaxies in \( V_i \). Therefore, the noise inflicted by the inaccurate approximation of the response by the events must be included in the data model.

The Poissonian PDF describes the number of events when the number of possible outcomes of an experiment are vast, but only a few counts are expected. This applies to the case for the observed number of galaxies in a box \( V \) where we have only the information of the expected number of enclosed galaxies. Of course, it would be desirable to include more knowledge about the environment of a galaxy into the PDF for their abundance, as many semi-analytical models do (e.g. White & Frenk 1991; Kauffmann, White & Guiderdoni 1993; Efstathiou 2000; Cole et al. 2000). Many N-body simulations indicate that there are numerous parameters that could or should be taken into account (e.g. Navarro, Frenk & White...
1996; Springel et al. 2005; Jenkins et al. 2001). However, adding more complexity makes the problem even harder to tackle than the Poissonian PDF. Besides, because of the generic nature of the Poissonian PDF, the method can also be applied to a completely different problem class, such as the shot noise of X-ray photons in a detector.

We now make the assumption that $d_i$, the number of events in $V_i$, is independent in its noise properties from all other data bins. In these assumptions, we follow the works of Layzer (1956), Peebles (1980), Martínez & Saar (2002) and Kitaura et al. (2009). Thus, the likelihood of the data becomes

$$P(d|\mu) = \prod_i \frac{\mu_i^{d_i} e^{-\mu_i}}{d_i!}. \quad (13)$$

As $\mu$ can be expressed in terms of $s$, this provides the desired likelihood $P(d|s)$. The data $d$ can also be viewed as the Poissonian sampled response. While $\mu$ itself is by definition noiseless, every realization $d$ has Poissonian noise.

3.1.2.3 Distortion operator. So far, the response (12) is entirely local. We now propose the concept of a distortion operator, which naturally introduces non-locality in the response (12).

As it is the data space that is distorted, we should apply the distortion to the local response (12). Hence, we define the distorted response

$$\mu_i = \sum_j R_{ij} \mu_j,$$

with a distortion matrix $R$.

It is a reasonable requirement that the distortion matrix should preserve the expected number of events, i.e.

$$\sum_i \mu_i = \sum_i \mu_i,$$

(15)
The most natural explanation of $R$ is to interpret $R_{ij}$ as the probability that an event which occurred at location $x_i$ in signal space is observed at location $x_j$ in data space and increases $d_i$. This is also how we set up $R$ in practice and it is easy to show that a distortion set up in this way fulfils the ‘conservation of $\mu$’ (15). Note that the factor $\kappa$ can be absorbed into $R$, which we assume from now on. Then, $R_{ij}$ has the meaning of the rate of events in data bin $i$ per $e^{h_{bs}}$ in signal space.

It is now in order to summarize in brief our data model.

(i) The log-density signal is a Gaussian field with covariance matrix $\mathbf{S}$, such that

$$P(s) = \frac{1}{\sqrt{2\pi |\mathbf{S}|}} e^{-s^\top \mathbf{S}^{-1} s/2}. $$

The matter distribution is modelled by a lognormal PDF, such that the density in $V_i$ is $\rho_i = \rho e^{\mu}.$

(ii) The response is given by $\mu_i = \sum_j R_{ij} \mu_j.$

(iii) The data likelihood given the signal is

$$P(d|s) = \prod_i \frac{\mu_i^{d_i} e^{-\mu_i}}{d_i!}. $$

It is evident that the complexity of this system strongly depends on the structure of $R$, and in particular whether $R$ itself depends on the signal $s$ to be reconstructed, or not.

The local data model with $R = \kappa^2 \mathbf{1}$ has been solved in a fully Bayesian framework by Enßlin et al. (2009). Implementation thereof in the field of large-scale structure reconstruction using the MAP method and a Hamiltonian sampler can be found in Kitaura et al. (2009, 2010), Jasche et al. (2010) and Jasche & Kitaura (2010). Of course, there are also other survey-based reconstructions of the large-scale structure, as can be found in Bertschinger et al. (1990), Yahil et al. (1991), Shaya, Peebles & Tully (1995), Branchini, Plionis & Scrima (1996), Webster, Lahav & Fisher (1997), Schmidt et al. (1999), Nusser & Hachmelt (1999), Hoffman & Zaroubi (2000), Goldberg (2001), Erdogdu et al. (2004, 2006), Vogelezang et al. (2004), Huchra et al. (2005) and Percival (2005).

3.2 Setting up a reconstruction problem

In order to test the map-making techniques on the lognormal density field with Poissonian noise, we generate mock data from a known signal and try to reconstruct the signal. Aiming at a proof of concept, it is in order to keep the complexity of the numerical simulation low, and hence we restrain ourselves to a 1D case. This restriction is not fundamental, because the same algorithm can also be applied to higher dimensional problems. In a sense, the topologies of the spaces involved are encoded in $S$ and $R$, but as we see below the inner structures of $S$ and $R$ are a priori irrelevant for our method on an abstract level. The only downside of multidimensional reconstructions is that the dimension of the field vector space grows as $N^2$, where $N$ is the number of pixels along any direction. So even moderate resolutions have a severe impact on the performance of any calculation and demand extensive optimization. Therefore, we content ourselves with $N$ pixels in one dimension for this conceptual work.

In order to be free of boundary effects, we choose a ring-like topology for our reconstruction, such that the first and the last pixels are direct neighbors. This ring-like topology is also not necessary as boundary effects can be treated naturally, but it simplifies the set-up.

The first step is to choose an adequate power spectrum for the signal. As we are mostly aiming for the principles of reconstructions with spatially distorted data, the details of the power spectrum are not essential here as long as most of the power is concentrated on large scales. Inspired by the Yukawa potential from quantum electrodynamics (Peskin & Schroeder 1995), which mediates a force with a range of $l_c$ or, in a different language, introduces correlation with a characteristic length-scale $l_c$, we try

$$P(k) \propto \frac{1}{k^2 + l_c^2}. $$

However, in order to have a more structured looking signal, we introduce a boost term, which selectively amplifies power on large scales:

$$P(k) \propto \frac{0.2 + e^{-0.008 k}}{k^2 + l_c^2}. \quad (16)$$

All reconstructions in this work were carried out for a correlation length $l_c = 0.05L$, where $L$ is the length of the simulation interval.

As the power spectrum is the diagonal of the Fourier transform of the covariance matrix, this allows us to easily compute

$$S(x) = \int dk e^{2\pi i (x-k) \cdot k} \tilde{P}(k) = F^{-1}[\tilde{P}](x - y). \quad (17)$$

1 The authors of this work use a mathematically similar response $\mu = \kappa [1 + B(e^{1-t} - 1)]$, which could also be used to model spatial distortion. Unfortunately, the authors do not mention the interconnection between their general scale-dependent bias $B$ and a PSF.

2 It may, however, have an impact on convergence speed, but these details are left for further investigation.
Here, $\mathcal{F}^{-1}[f](x)$ must be read as the inverse Fourier transform of $f$ evaluated at $x$ and $\hat{P}(k)$ denotes the matrix with $P(k)$ on the diagonal. Similarly, we can compute
\begin{equation}
S_{\alpha}^{\alpha} = \mathcal{F}^{-1}[1/\hat{P}](x-y) \tag{18}
\end{equation}
and
\begin{equation}
T_{yy} = \mathcal{F}^{-1}[(\hat{P}^{1/2})(x-y)], \tag{19}
\end{equation}
which we need for the evaluation of $H_d$ and to generate mock signals, respectively. In Appendix B, we describe how a random signal with given covariance matrix $S$ can be constructed from a vector of random numbers using $T$, the root of $S$.

### 3.2.1 Naïve data inversion

As a competitor for the fully Bayesian reconstruction, we also construct an unoptimized filter, namely a non-Bayesian reconstruction, which we call $m_{na}$. The naïve data inversion is not as straightforward as it may seem at first sight. Although the performance of $m_{na}$ is extremely poor it is still worthwhile investigating the problems of direct data inversion in order to understand why this approach is doomed to fail.

In principle, naïve data inversion seems sensible, as the data $d$ trace the response $\mu$, which can directly be inverted to yield $s$. However, $d$ is not exactly $\mu$ because of the Poissonian sampling of the response. As a non-Bayesian analysis neglects the signal covariance for $m_{na}$, no information on the true underlying $\mu$ can be inferred from adjacent pixels. Hence, there is no way to tell if the data point $d_i$ is lower or higher than the expected $\mu_i$, and the only choice is to make the approximation $\mu_i \approx d_i$. Furthermore, when the distortion depends on the signal, there is no way to consistently incorporate bin mixing in the data inversion, such that we have to approximate the distorted response by the local response (12). This leads to
\begin{equation}
\tilde{m}_{na}(d_i) = \frac{1}{b} \log \left( \frac{d_i}{\kappa_i} \right). \tag{20}
\end{equation}
However, this is only possible for pixels with $d_i > 0$, and a generic guess such as $\tilde{m}_{na}(d_i = 0) = 0$ for these pixels cannot be correct, as for high galaxy densities with $w_i V_i P_{00}^m = \kappa_i > 1$ the estimated signal is negative even for $d_i = 1$. A way out is to allow for Bayesian reasoning in this case and to ask for the most probable field strength $\kappa_i$ given its variance and $d_i = 0$. This leads to
\begin{equation}
\tilde{m}_{na,i} = \left\{ \begin{array}{ll}
(1/b) \log(d_i/\kappa_i) & d_i > 0 \\
-(1/b) W(w b^2 S_{ii}) & d_i = 0,
\end{array} \right. \tag{21}
\end{equation}
where $W(z)$ is the Lambert W-function (i.e. the root of $z = w e^w$). As it stands, $\tilde{m}_{na}$ is however a bad estimator for the signal as the noise from the Poissonian sampling is very present and the true signal is known to be smooth.

It is common practice to apply a smoothing procedure to the inverted data. However, the shape of the smoothing kernel introduces new free parameters and there is no generic setting for this in non-Bayesian reasoning. Allowing for the use of the covariance $S$ again gives us an alternative, as it can be used to smooth the inverted data. We choose to convolve with $T$, the root of $S$, as it is more localized and therefore the convolved map reflects the data better. Hence, the final non-Bayesian map is given by
\begin{equation}
m_{na} = T \tilde{m}_{na}. \tag{22}
\end{equation}
Even though $m_{na}$ gives really poor guesses for the underlying signal $s$, our tests have shown that smoothing with a top-hat filter performs even worse. This also justifies our choice of $T$ as a smoothing kernel.

We would also like to stress that even this naïve map construction had to rely on some Bayesian elements that signal prior information was necessary for treating the $d_i = 0$ case and setting up optimal smoothing. Besides, at a closer look the smoothing procedure is questionable, because it introduces a hidden prior for a signal with a certain correlation length or power spectrum. However, instead of introducing a hidden prior, we should rather be upright, make clear where the power spectrum for the signal enters the reasoning, and include the prior for $s$ right from the start, as we propose in the next section.

### 3.2.2 Bayesian map-making

Because a full posterior mean, as proposed in Section 2.2, is not straightforward for this problem, we have to rely on approximations for $(\langle s \rangle_i | d_i)$. One possibility would be to sample the posterior via a Monte Carlo Markov chain (MCMC) method. However, this is computationally very expensive and difficult to understand analytically, and therefore not suited for the proof of concept at which we aim. Instead, we use the approximation of $(\langle s \rangle_i | d_i)$ by the classical map, as proposed in Section 2.4.

For this, it is sufficient to minimize the probability Hamiltonian of the problem defined by equation (5). For our problem defined by the likelihood (13) and the signal prior (7), it is given by
\begin{equation}
H_d[s] = -\log P(d | s) = -\log [P(d | s) P(s)] = {1 \over 2} s^T S^{-1} s - \sum_i d_i \log \mu_i + \sum_i \log (\prod \mu_i). \tag{23}
\end{equation}
where we have shifted the Hamiltonian by a constant in the last step.

As the classical map $m_d$ aims at minimizing the Hamiltonian, we can use efficient multidimensional minimization techniques. In particular, we use the conjugate gradient method in the Broyden–Fletcher–Goldfarb–Shanno variant, as provided by the GNU Scientific Library to minimize $H_d[s]$.

The conjugate gradient method needs the derivative of the function to be minimized, so we calculate
\begin{equation}
\delta H_d / \delta s = S^{-1} s - \left( \frac{d}{\mu} - 1 \right)^T \frac{\delta \mu}{\delta s}. \tag{24}
\end{equation}
For an introduction to the method of conjugate gradients, see Shewchuk (1994).

6 The GNU Scientific Library is available from [http://www.gnu.org/software/gsl](http://www.gnu.org/software/gsl).
When $R$ is independent of $s$, the gradient of $\mu$ is easily computed:

$$\frac{\delta \mu_i}{\delta s_k} = R_{ik} b e^{h_{ik}}.$$  \hfill (25)

### 3.2.3 Error estimation

The signal reconstruction alone is only of little use, as it contains no information on the reliability of the reconstruction. An approximation of the correct error intervals for $m_i$ can be obtained via approximation of the posterior with a Gaussian. Therefore, we Taylor-expand the probability Hamiltonian up to second order in $s$ and obtain

$$H[s] \approx H_0 + \frac{1}{2} (s - m_d) ^ {\top} D ^ {-1} (s - m_d).$$

The Gaussian approximation of the posterior is then given by

$$P(s|m) \approx G(s-m_d, D).$$

With the generating functional from equation (8) it is then easy to calculate the expected deviation from $m_i$

$$\langle \delta s_i, \delta s_j \rangle_{\theta(s|m)} = \frac{\delta^2}{\delta J_i \delta J_j} \log Z_{|J=0} = D_{s_j},$$  \hfill (26)

where $\delta s = s - m_d$. We should keep in mind that this procedure gives only an approximation of the real 1σ confidence interval because of the Gaussian approximation of the posterior and also does not display the cross-correlation between errors at different locations. Visual inspection of the error bars obtained for our example figures shows that they are neither too large nor too small and are therefore a good approximation of the correct error bars.

In our case where the Hamiltonian is given by equation (23) we find

$$D_{\mu m} ^ {-1} = S_{\mu m} ^ {-1} + \sum_i \left( \frac{d}{\mu_i} \delta \mu_i \frac{d}{\delta s_m} - \sum_i \delta \mu_i \right),$$  \hfill (27)

where we have not implied any summation for clarity.

If $R$ depends on the signal, we have to make the simplifying assumption that it varies only slowly with $s$ such that its contribution to the gradient of $\mu$ can be neglected. The reason for this simplification is mainly that for our specific $R(s)$ model developed in Section 3.6.1 this derivative is computationally extremely expensive to calculate, let alone the second derivative of $R$. Therefore, we obtain

$$\frac{\delta \mu_i}{\delta s_m} = b R_{im} e^{h_{im}}, \quad \frac{\delta^2 \mu_i}{\delta s_m \delta s_n} = \delta_{mn} b^2 R_{im} e^{h_{im}}. $$

which can be used in equation (27).

### 3.3 Photometric redshifts

We apply our reconstruction algorithm to different cases, one of which is the reconstruction of dark matter density from galaxy counts whose redshift is determined photometrically. Photometric redshift measurement is based on the apparent colours $c$ of galaxies rather than their full spectra, and schemes to assign a redshift $z$ depending on the measured $c$. However, as Benitez (2000) points out, there is no unambiguous colour to redshift mapping even if more and more band filters are used. He argues that instead of a strict mapping $z = z(c)$, we have to work with the full posterior $P(z|c)$ that a galaxy with colour $c$ actually has redshift $z$.

Wittman (2009) treats this problem by drawing a random $z_{MC}$ from $P(z|c)$ for each galaxy and continues his calculation with this $z_{MC}$. However, he is aware that this procedure only works if the number of galaxies per colour bin is large. So we choose a density field reconstruction from photometric redshift data as a testing ground for our method.

#### 3.3.1 Distortion matrix for photometric redshifts

The distortion matrix for photometric redshift distortions maps from redshift into colour space and must be set up according to $R_{cz} = P(c|z)$. The features we want to test for are the asymmetric shape of $R$ and the robustness to catastrophic outliers. As this is intended as a test case, we make some simplifying assumptions about $P(c|z)$. First, we assume that the shape of $P(c|z)$ remains fixed when going to higher redshift and secondly we neglect the effect of the spectral type on the colour PDF. Catastrophic outliers are modelled by a small Gaussian PDF contribution, which is offset from the main peak by half of the simulated interval length and whose height was chosen to be one-tenth of the main peak. It should be stressed, however, that once $R$ is set up in a realistic way, no change is needed for the algorithm. Fig. 1 shows the $P(c|z)$ we use; it should be explicitly said that this is not a real-world $P(c|z)$ but one that we designed to just look similar to a real-world $P(c|z)$, as in Benitez (2000).

#### 3.3.2 Reconstruction of redshift space matter distribution from colour space data

For our reconstruction test, we set up our simulation on an interval of length $L = 1$ split into 1024 evenly sized pixels. As we are not interested in boundary effects here, we set the window function to unity for the whole interval.

Figs 2 and 3 show reconstructions of the signals with unit variance $\langle s^2_i(z) \rangle = 1$ from photometrically distorted data. Fig. 3 also shows the characteristic residuals for this single signal $s_1$ when reconstructed

![Figure 1. The probability distribution of observed colour $c$ for three different redshifts. The $x$-axis is the number of the colour bin the galaxy may be put into. Here, we use 1024 colour and redshift bins and the maximal redshift bin corresponds to $z = 1$. Note that the shape of the distribution is assumed to be independent of $z$ and is only shifted towards higher colour values. This is a simplification with respect to real photometric redshift probabilities, where the shape of the PDF does depend on $z$. As we use a cyclic topology for this test case, $z = 1$ corresponds to $z = 0$, and $c = 0$ to $c = 1024$. Therefore, the catastrophic outliers for $z = 0.500$ (dotted red) reappear at the low colour values on the left.](https://academic.oup.com/mnras/article-abstract/409/4/1393/984616)
by this technique, defined as

$$\text{Res}^{(n)} = \left( m_d - s \right)^2 \frac{1}{d(n)},$$

where $m_d$ is the map as it is reconstructed from the data $d$. In practice, we perform this average for 500 different data realizations of the same signal $s$. This shows in which regions the algorithm generally performs badly for a specific signal and where it does well independently from the data realization. It can be seen in Fig. 3 that in actual data realizations the galaxy numbers vary for different biases while keeping $\bar{\rho}_{gal}$ constant. This has to be expected because the average galaxy density is only the same for different biases, when averaged over all signal realizations. This may affect the comparability of the characteristic residuals for different biases, but not for the same bias and different galaxy densities.

The first thing to notice about the reconstructions is that in all cases they are smooth even though the noise in the data is considerable and also the original signal has far more power on small scales. This is against common knowledge that MAP maps pick up a lot of noise. What prevents this in our case is the distortion matrix, which acts as a smoothing operation on the observed data such that $m_{cl}$ itself becomes smooth again.

In Fig. 2, we also show a classical reconstruction $m_{cl}^1$ that does not take the distortion of data space into account (i.e. it assumes that $R = \kappa \mathbf{1}$). Compared to our full reconstruction $m_{cl}$, we can see that neglecting the colour space distortion results in severe difficulties to get even the shape right. Instead of voids, $m_{cl}^1$ even sees peaks at 0.4 and 0.65. Although large peaks are detected, their height is slightly underestimated because some events are scattered away, which $m_{cl}$ is not aware of. The full reconstruction, however, can even reshape the void region located at 0.4, despite the abundance of events in that region originating from the large peak at 0.9. Looking at the characteristic residuals in Fig. 3 reveals that this is true for most data realizations.

In Fig. 3 we show reconstructions of the same signal for different simulation parameters and their average residuals. In the reconstructions, we can make out two obvious trends for our simplified model:

(i) the higher the galaxy density, the better the overall reconstruction – note in particular the decreasing level of $\text{Res}^{(n)}$;

(ii) the higher the bias, the better the reconstruction of peaks, but the worse the reconstruction of void regions.

Both observations are not surprising, as higher galaxy density means that the response is sampled with less Poissonian noise, so we expect the reconstruction to become better. Higher bias, however, sharpens the contrast such that peaks are selectively sampled with high accuracy, whereas the galaxy density in void regions is reduced. This trend is also reflected by the error bars, which tighten up in overdense regions for biases larger than 1, but not so for $b = 0.5$.

The example for $b = 0.5$ and $\bar{\rho}_{gal} = 500$ shows that data variance can make a big difference. Although the number of galaxies is nearly twice as large as in the panel to its left, the peak at 0.55 is hardly detected at all. Looking at the residuals reveals that this is simply an artefact of a lucky data realization for the low-density case and an unlucky one for the middle-density case.

Looking at the characteristic residuals, we notice that for $b = 0.5$ the general trend of the signal (i.e. the largest Fourier modes) can be detected even for very low galaxy densities, while sharp peaks and ditches are even difficult to resolve for $\bar{\rho}_{gal} = 1000$. For $b = 1.5$ and $b = 2.5$, the overdense regions can be resolved even for a low density of events, while the voids are difficult to reshape for low galaxy densities.

It should be mentioned, however, that the poor resolution of voids is also a problem of this specific example, because every void region has an overdense region that scatters events into the void (0.9 into 0.4 and 0.15 into 0.65). Had we chosen an example where two void regions scatter into each other, the resolution of these voids would be much better.

Table 1 lists the average residuals of 500 reconstructions of different signals. The naive map $m_{cl}$ is in all cases the worst reconstruction. Especially for low galaxy densities it is painfully close to the average residual of a zero map, which is expected to be unity for a Gaussian signal with unit variance. This tells above all one thing: for a reconstruction in the low signal-to-noise regime we must include some knowledge about the signal. For a low bias $m_{cl}^1$ and $m_{cl}$ have roughly the same error level when the galaxy density is low. When $\bar{\rho}_{gal}$ rises, both maps become better but what surprises is that $m_{cl}$ seems to saturate over $\bar{\rho}_{gal} \geq 500$ at an error level of ≈0.15. Considering the fact that the distortion of the signal inevitably destroys some information, the question arises whether this is already the optimum. Therefore, we let the same signals be processed without noise in the response (i.e. with perfect data) and we have found for $b = 0.5$ that without Poissonian noise $m_{cl}$ has an average residual of 0.09 and $m_{cl}$ of 0.27. So we see that for the signals with $\bar{\rho}_{gal} = 500$ and $\bar{\rho}_{gal} = 1000$ there is still some margin to be gained by even higher galaxy densities, but not much. It is remarkable that $m_{na}$ with noiseless data comes not even close to the performance of $m_{cl}$ with medium galaxy density.

There is one anomaly worth noticing, namely the increase of the residuals from $b = 1.5$ to $b = 2.5$ observed for all maps but $m_{cl}^1$ in particular, which comes from catastrophic outliers scattering into void regions. As count rates for $b = 2.5$ are only large at big overdensities and remain small even for smaller peaks, $m_{cl}^1$ makes a small peak from scattered events in voids. The same holds also for $m_{na}$, but this map does not react as quickly on few events as $m_{cl}$ does, which explains the modest deterioration of $m_{na}$ and the striking difference for $m_{cl}^1$. 

© 2010 The Authors. Journal compilation © 2010 RAS, MNRAS 409, 1393–1411
Figure 3. The top parts of the nine panels show reconstructions of the same signal (dotted red) from data (grey dots) with Poissonian noise and typical spatial distortions as they occur in photometric redshift measurement. The data are where the point-wise inversion from equation (21) would place them. The blue line shows the classical signal reconstruction as proposed by our method with its $1\sigma$ error contours (thin blue lines). The bottom parts of the nine panels show the characteristic residuals as defined by equation (28) for 500 different data realizations for this signal.
3.4 Signal inference from independent data sets

In real-world surveys such as SDSS there might be both available: accurate spectroscopic galaxy redshift and abundant less certain photometric redshift data. Can the former help to better localize the latter?

Our formalism allows us to combine both data sets by defining

\[ d = [d^{(1)}, d^{(2)}] = [R^{(1)}, R^{(2)}] \kappa \, e^{h_s}, \]

(29)

and business is as usual.

In principle, this allows us to combine any two data sets. However, now we consider the case where \( d^{(1)} \) are data from photometric redshift measurement and \( d^{(2)} \) are data from spectroscopic redshift measurement. As such, we model \( R^{(1)} \) as in Section 3.3.1. Because spectroscopic redshift measurement is far more accurate than photometric redshift measurement, we assume spectroscopic redshift measurements to be exact, that is, \( R^{(2)} = 1 \kappa \), with \( \kappa \) being the zero-response (12).

Fig. 4 shows a reconstruction for one data realization \( d^{(1)} \) and \( d^{(2)} \) with a bias \( b = 1.5 \). The combined reconstruction \( m^{cl+(ph)} \) mostly follows the reconstruction from spectroscopic data \( m^{cl}_{gal} \), especially in overdense regions such as from 0.8 to 1.0 and 0.5 to 0.6. This is an expected behaviour, as in this case \( b = 1.5 \) and therefore overdense regions are accurately sampled even with the low average galaxy density \( \bar{\rho}_{gal} = 60 \). Also, because the spectroscopic data have no spatial error, they dominate the reconstruction in this regime. However, in regions where spectroscopic galaxies are rare, the combined reconstruction sometimes deviates substantially from \( m^{cl}_{gal} \) (most conspicuous from 0.2 to 0.4). The crucial point to note is that the combined reconstruction is not an average of the two reconstructions \( m^{cl}_{gal} \) and \( m^{cl}_{ph} \) as can be seen at 0.4 to 0.5 and 0.6 to 0.7 where the combined reconstruction lies below the others. Therefore, the optimal combination of the two data sets is non-trivial and non-linear.

The statistics for 500 different signal realizations (not shown) confirm that the combination is more than just a superposition of the two single-data-set reconstructions.

3.5 X-ray astronomy via coded mask telescopes

Thanks to the generic structure of our approach, we can also apply it to a completely different problem. One such problem is the detection of extended sources in coded mask aperture systems.

Coded aperture systems were originally proposed by Dicke (1968) for the purpose of detecting point-like X-ray sources. In this scheme, an absorbing plate with a pattern of pinholes is placed in front of a detector and the shadow of this plate on the detector allows us to unfold the count distribution and infer the positions of the X-ray sources. However, this technique becomes much more difficult when the light sources are extended and not just point sources.

We now demonstrate that it is, in principle, possible to map out extended sources with our method, when the count rate and bias are high enough. Adapting our method to this problem, the mixing matrix in this case must be the coded mask pattern that lets light pass for open pixels and shields it completely for dark pixels. Hence, we only need the pattern of a coded mask for our distortion matrix. Today’s coded masks have an optimized pattern, but for our purpose the originally proposed random pattern (Dicke 1968) of blocking and transparent pixels is sufficient.

Fig. 5 shows two reconstructions on an interval with 512 pixels from data obtained via a random pattern coded mask. In this set-up, it is only possible to detect the largest peak and therefore we use a large bias of \( b = 2.5 \). It is remarkable that a surprisingly low number of photons \( n_{ph} = 404 \) are sufficient to detect the peak at position 0.25 in the top panel and even some parts of its substructure. In regions where the algorithm cannot see the signal, the error bars widens up to the interval +1 to −1. This is a good consistency check, because this is expected for a signal with unit variance when no information on the signal is available.

The lower panel in Fig. 5 shows a reconstruction from a signal with an extremely high peak. Although such high peaks and count rates are rare, it is interesting how many details of the peak can be reconstructed with an extremely small error range. Note that the smaller peak at 0.15 in this example (although comparable in size to the peak in the top panel) is not detected at all. This is because of the overwhelmingly larger brightness of the largest peak at 0.75 whose photons hit the same detector, but with an approximately \( e^{1/(1+C)^2} \approx 148 \) higher rate. The Poissonian noise from the larger peak overlaps with the photons from the smaller peak, rendering it impossible to detect.

This is however not yet a fully realistic set-up and further effort must be put into refining this technique, as sometimes false peaks (with narrow error bars) turn up in the reconstruction. It is possible that this problem has already been amened by the invention of
Reconstruction of spatially distorted fields

Figure 5. Reconstruction of a signal (red dotted line) from data (grey dots) that mimic the behaviour of coded mask apertures with a bias $b = 2.5$ and an average photon count rate $n_{\text{phot}} = 500$. The $m_{\text{cl}}$ reconstruction (blue smooth curve) shows only the largest peak of the signal, and some of its substructure. $1\sigma$ error levels are indicated by thin lines. Being completely useless, the na"ive map $m_{\text{na}}$ was omitted.

3.6 Reconstruction with s-dependent distortion

We now address the problem where the distortion matrix depends on the signal to be reconstructed. The paradigm for this is the measurement of distance via redshift. Because of the peculiar velocities of galaxies with respect to the Hubble frame, the comoving distance of galaxies cannot exactly be calculated from their redshift. Note that there does not need to be a coordinate transformation from redshift into real space, because objects with different positions may have the same redshift. Therefore, the mapping from real to redshift space is not injective, and thus not invertible.

In particular, the gravitational pull of matter overdensities affects the peculiar velocities of galaxies. So if we want to unfold the large-scale matter distribution in the Universe by measuring angular positions and redshifts of galaxies, we have to take the distorting effects of matter on the redshift space into account. What makes this problem so particularly demanding is that the distortion operator, which transforms the real-space matter distribution into the observed redshift space matter distribution, depends on the real-space matter distribution itself which is to be reconstructed.

We have to be aware that the assumption that the matter field is sampled by a Poissonian process will fail at some point when going to ever smaller scales. Here we show the problems we have to face even in an idealistic set-up where the forward transformation from real to redshift space is perfectly known.

3.6.1 A statistical model for redshift space distortions

We now address the forward problem of constructing the redshift space galaxy distribution from a known real space matter distri-

...
system is given by a Boltzmann factor

\[ P(v_{nl}) \propto \exp \left( -\frac{v_{nl}^2}{\Phi} \right), \quad (32) \]

which guarantees the right velocity dispersion. The fact that non-linear velocity distortions are given by a Maxwellian distribution is a common assumption used by many authors (e.g. Peacock & Dodds 1994; Heavens & Taylor 1995; Tadros et al. 1999; Tadros & Efstathiou 1996). The other school of thought models the non-linear velocity distribution as an exponential pairwise PDF, i.e.

\[ P(v) = (2^{1/2} \sigma)^{-1} \exp \left( -2^{1/2} |v|/\sigma \right) \]

(e.g. Bromley, Warren & Zurek 1997; Hamilton 1995; Ratcliffe et al. 1998).

We now have to blend the two regimes into one general expression. Therefore, we extend equation (32) and make the ansatz

\[ P(v_{ij}) \propto \exp \left[ -\frac{(v_{ij} - v_{ij,0})^2}{2F(\Phi)} \right], \quad (33) \]

where \( F \) is a continuous function to be determined and \( v_{ij,0} \) is the component along the line of sight of the linear velocity field determined by equation (30). The function \( F \) must have the following limiting behaviour in order to interpolate between the linear and non-linear regimes:

\[ \lim_{\Phi \to +\infty} F(\Phi) = |\Phi|/2 \quad \lim_{\Phi \to -\infty} F(\Phi) = 0. \]

As \( F \) stands in the denominator, it must never become exactly 0, and hence it must also not change sign. Apart from this, the Boltzmann factor should play little to no role in the linear regime where \( \Phi \) is above a threshold \( \Phi_0 \) over which we do not assume objects to be virialized. In principle, we should therefore require that \( F \approx 0 \) in these regions (i.e. exact measurement). In practice, however, every measurement comes with uncertainties, and even in the linear regime galaxies can have unpredictable small peculiar velocities. Subsuming this as ‘measurement uncertainties’, it makes sense to require that \( F(\Phi) > \sigma_0 \), which includes this instrument noise naturally in our formalism. In other words, we turn the shortcoming of our method to model the exact measurement into the feature to always include a minimal error with variance \( \sigma_0 \). For the non-linear regime, we assume that galaxies are only virialized with their potential excess \( \Delta \Phi = \Phi - \Phi_0 \) [i.e. \( F(\Delta \Phi) = -\Delta \Phi/2 \)].

One possibility to smooth the transition from linearly to non-linearly dominated regions is by a tilted hyperbola such as

\[ F(\Phi) = \frac{1}{4} \left[ \sqrt{(\Delta \Phi + 2\sigma_0^2)^2 + \tau^2} - \Delta \Phi \right] + \frac{1}{2} \sigma_0^2. \quad (34) \]

This has the advantage of only introducing one more free parameter \( \tau \), which controls the smoothness of the transition.

This allows us to write the distortion matrix for the transformation from real to redshift space as

\[ R_{ij} = P(z_i|x_j, \rho) \propto \exp \left[ -\frac{(z_i - v_{ij,0})^2}{2F(\Phi_j - \Phi_0)} \right]. \quad (35) \]

Altogether we now have five free parameters to control the behaviour of our self-built velocity dispersion model:

(i) the strength of the linear velocity distortions, which is given for a cosmological model;

(ii) the width of the top-hat low-pass filter for the linear velocity field, which can be determined by analysing the redshift space matter power spectrum as \( k \lesssim 0.15 \, h \, \text{Mpc}^{-1} \) (Percival et al. 2007; Smith et al. 2003);

(iii) \( \tau \), adjusting the smoothness of the transition from linear to non-linear behaviour;

(iv) \( \Phi_0 \), for the zero-level of the gravitational potential under which non-linear effects kick in;

(v) \( \sigma_0 \), for any further measurement error.

As we are aiming at a proof of concept, we also take the gravitational constant, which adjusts the strength of the non-linear effects, as a free parameter to generate visible effects with our distortion model. Fig. 6 shows the mixing matrix as it was used in the reconstructions of the following section.

As before, we use the conjugate gradient method for minimization of the probability Hamiltonian. This method requires the derivative of the Hamiltonian and therefore also the derivative of \( R[z] \) with respect to \( z \). However, the computation of \( \delta R/\delta z \) is not feasible for our model, as it would demand the calculation of \( N_{\text{pix}}^3 \) expensively to compute values. This would severely deteriorate the performance of our algorithm, as the evaluation of \( R \) already is the bottleneck. Therefore, we make the approximation that the contribution of \( \delta R/\delta z \) is small compared to the other terms in

\[ F(\Omega) = \frac{1}{2} \left( \frac{2f(\Omega)/3aH}{2f(\Omega)/3aH} \right) \approx 1.28, \quad \tau = 5 \times 10^{-4}, \quad \sigma_0 = 1.75 \times 10^{-4}, \quad \Phi_0 = 0 \]

and the top-hat filter lets the lowest 1 per cent of \( k \)-modes pass.

---

7 For completeness, we give the numerical values of our settings: \( 4\pi G a^2 = 0.25, [2f(\Omega)/3aH^2] = 1.28, \tau = 5 \times 10^{-4}, \sigma_0 = 1.75 \times 10^{-4}, \Phi_0 = 0 \) and the top-hat filter lets the lowest 1 per cent of \( k \)-modes pass.
3.6.2 Results from the reconstruction of the real space matter distribution from redshift space data

We set up a run of 500 signal reconstructions on an interval of length \( L = 1 \) with 1024 equally sized pixels. Because of our limited interval length, we had to restrict ourselves to signals whose maximum peak was less than 3.5, because for peaks higher than this value the tail of the smeared-out peaks overlaps with the tail on the other side.\(^8\)

For each signal realization \( s \) we generate one possible data realization \( d \) and perform three different types of reconstruction, namely the naive map \( m_{\text{na}} \) from Section 3.2.1, a MAP reconstruction neglecting redshift space distortions \( m_{\text{cl}}^0 \) and the full MAP map including the model redshift space distortions \( m_{\text{cl}} \). Fig. 7 shows our \( m_{\text{cl}} \) reconstruction for one signal but for different \( b \) and \( \tilde{\rho}_{\text{gal}} \) settings, along with the characteristic residuals \( \text{Res}^{(s)} \) for the reconstruction of the signal in the lower panels. For our model we find the following general trends:

(i) the higher the galaxy density, the better the reconstruction;
(ii) the higher the bias, the worse the reconstruction of voids, but peaks however are better reconstructed.

The reasons for these trends are the same as in Section 3.3.2. The reconstruction from galaxies with redshift distortions is therefore similar in many respects to that with photometric redshifts.

Note that the centre peak in the data (at 0.55) shows a clear dislocation from its signal counterpart in the redshift space of the large peak at 0.9, an effect resulting from the linear velocity distortions. The \( m_{\text{cl}} \) reconstruction, however, being aware of the redshift distortions introduced by the large peak, places the peak at the right spot in all reconstructions. Interestingly, for large biases this peak is well resolved even for very low galaxy densities, as the characteristic residuals show.

Yet there are limitations for the reconstructions, as can be seen in the indentations flanking the largest peak at positions 0.8 and 1.0. Although being comparable in size to the peak at 0.55, they are not resolved in any reconstruction, and this even remains the case for extremely large \( \tilde{\rho}_{\text{gal}} \), which we do not show here. Apparently, these structures are irreversibly lost in the shadow that the larger peak at 0.9 casts on close-by smaller structures. This is also not unexpected, as the distortion matrix \( R \) is set up in a way that large peaks become smeared out over large distances, whereas small peaks remain localized. Therefore, the smaller peak appears as an extension of the plateau. The most important characteristic of this kind of error is that it does not improve with higher \( \tilde{\rho}_{\text{gal}} \) in contrast to areas where more information can be gained by reducing the noise (e.g. wide void regions).

Fig. 8 shows the centre panel from Fig. 7 with the reconstruction for \( b = 1.5 \) and \( \tilde{\rho}_{\text{gal}} = 500 \) again but also the naive map \( m_{\text{na}} \) and the MAP map neglecting redshift space distortions \( m_{\text{cl}}^0 \). At first sight, we notice the poor guess that \( m_{\text{na}} \) gives, which we take as an argument that Bayesian analysis is inevitable to tackle this problem.

The more interesting competitor for the MAP map, including the correction of redshift space distortions \( m_{\text{cl}} \), is \( m_{\text{cl}}^0 \). In general, the shape of the reconstruction is very similar, which is as it should be, as both algorithms are based on the same principles and work on the same data set. Yet there are some distinct features that make a difference. Most prominent is, of course, the correct treatment of the linear redshift space distortions of \( m_{\text{cl}} \) in contrast to \( m_{\text{cl}}^0 \), which sees the void from 0.15 to 0.3 further depleted and the peak at 0.55 displaced towards the massive on the right. Another more subtle difference is that \( m_{\text{cl}}^0 \) picks up more small-scale features from the data, as can be seen in region 0.8–1.1. This is because of the smoothing effect of \( R \) on the \( m_{\text{cl}} \) reconstruction, as already discussed in Section 3.3.2.

In contrast to the photometric case from Section 3.3.2, the error bars do not tighten up in overdense regions. However, this is to be expected, as our model was set up in such a way that the position uncertainty in the neighbourhood of large peaks is largest. In particular, this makes the detection of substructure in large peaks nearly impossible.

For a signal-independent view on the reconstruction quality, we list in Table 2 the average \( L_2 \)-distance from reconstruction to true signal for 500 different signal reconstructions.

The reconstruction benefit from including redshift space distortions seems not to be overwhelming but tends to be larger if bias and galaxy density rise. Notably, for \( m_{\text{cl}}^0 \) the effects from linear and non-linear redshift distortions partially cancel each other for the parameters chosen here and thereby improve the performance of \( m_{\text{cl}}^0 \). This is because non-linear redshift space distortions smear out large peaks while linear redshift distortions compress large overdensities. If we turn off linear redshift distortions and set up \( R \) only with our virialization model for non-linear redshift distortions, we obtain the interesting effect that the average \( L_2 \)-distance of \( m_{\text{cl}} \) to \( s \) becomes smaller, but for \( m_{\text{cl}}^0 \) it increases instead.

So far, we have assumed that the underlying redshift distortion model is perfectly known and its parameters are the same for both data generation and the reconstruction phase. In a set-up for measured data, this is not the case. Neither is there a perfect model for the forward transformation from real to redshift space, nor are its parameters accurately measured. How severely parameter errors affect the reconstruction has not been scrutinized, but can be the aim of further investigation. Still, if this model was adapted and applied to measured data, the above-mentioned reconstruction characteristics and limitations would hold.

3.6.3 Refining the distortion model

In Section 3.6.1 we introduced a model for the transformation from real to redshift space in a statistical way. One important step was to apply a low-pass filter to the potential before the linear velocity field was calculated. However, similarly to using only the low modes of the perturbation for calculating the velocity distortions, it would be
Figure 7. The top parts of the nine panels show the real space reconstructions of the same signal (red dotted line) from data given in redshift space (grey dots) with Poissonian noise where the transformation from real to redshift space is given by equation (35). The data are where the point-wise inversion from equation (21) would place them. The smooth blue curve shows the classical signal reconstruction as proposed by our method with its 1σ error contours (thin blue lines). The bottom parts of the nine panels show the characteristic residuals as defined by equation (28) for 500 different data realizations of this signal.
reasonable to use only the high modes for calculating the velocity dispersion.

This is because the non-linear velocity dispersion predominantly depends on small-scale physics. If a halo with many galaxies collapses, the energy that the galaxies gain from the collapse is only the difference from their former potential energy and the potential energy after they have entered the collapsed structure. The velocity dispersion is hence not proportional to the full potential, but only to the fraction of the potential that the galaxies have actually fallen. Assuming that the galaxies in the collapsed structure are all from the vicinity of the collapsed object, this fraction can be estimated as the high-frequency component of the potential. In the picture of energy conservation we can look upon this in the following ways:

(i) the low-frequency part of the potential adds to the energy of linear velocity distortions;
(ii) the high-frequency part of the potential adds to the non-linear velocity component.

Fig. 9 shows the resulting distortion matrix of the modified model for the same signal as in Fig. 6.

When we set up the velocity distortions in this way, we find that the problem becomes much harder and the MAP solution ultimately fails to give reasonable results. Interestingly, it is the low bias case $b = 0.5$ that is hardest. This is most likely because the response for $b = 0.5$ makes a double peak from large peaks (i.e. bifurcates the peak). This misleads the MAP map to make two peaks out of one, which can lead to very weird behaviour. This bifurcation will eventually also happen for the larger biases, if we turn up the strength of the velocity dispersion.

Our tests show that the MAP solution via conjugate gradient minimization has severe problems with this bifurcation. We also employed a simulated annealing technique for minimization, which gave us the same results. So we can say, with confidence, that in the case of a bifurcated response the MAP method may give the most likely map, but still a very bad reconstruction. With the complexity of the distortion matrix at this point, we finally have reached a limit for the MAP method.

### 3.6.4 Comparison to Metropolis–Hastings sampling

The question now arises whether the MAP method is already the optimal reconstruction given the data or not. According to theory, this should not be the case because the MAP map minimizes $\langle \| m - s_t \|_2^2 \rangle_{\hat{m}_{gal}}$, but the $L_2$-norm is a distorted measure of distance. Unfortunately, the posterior is too complex that $(s_t)_{\hat{m}_{gal}}$ can be calculated directly. However, if there were a way to approximate the
Figure 10. On the left are two reconstructions for different settings of the biases $b = 0.5$ and $b = 1.5$. The signal appears as a red dotted line, the data (grey dots) are where the naive data inversion from equation (21) would place them. The naive map (thin evenly dashed line) is the smoothed data and shows signs of bifurcation at 0.9. The $m_{cl}$ reconstruction (blue dash-dotted curve) interprets the bifurcation as two different peaks, while the posterior mean (smooth green curve) does not. 1σ error contours are indicated by thin lines. The pictures on the right show the uncertainty covariance matrix of the Metropolis–Hastings reconstruction. Note the long-range influence of the structure at 0.9.

posterior by another PDF, which is easier to evaluate, we might be able to calculate $\langle s \rangle_{d|s}$ directly. As our interest is not in the detailed shape of the posterior, but our aim is simply to evaluate integrals over the posterior quickly, we may approximate the posterior $P(s|d)$ by

$$\tilde{P}(s) = \frac{1}{N} \sum_{i=1}^{N} \delta(s - s_i),$$

where we construct $s_i$ using a Metropolis–Hastings MCMC method. In principle, the chain can start from any map $s_0$, but for the sake of skipping the burn-in phase, we start our MCMC from $m_{cl}$. At any given point $s_n$, we generate a small variation $\delta s$ with the same power spectrum as the signal, but with far smaller amplitude and we set $s_{n+1} = s_n + \delta s$. We accept or reject this new sample according to the Metropolis–Hastings criterion (e.g. Binder 1997). This gives us a chain of maps where consecutive samples are correlated, but after some steps the correlation vanishes.

Because the MCMC is expensive with respect to computation time, we cannot run several hundred reconstructions but have to be satisfied with the evaluation of selected cases. Therefore, we show in Fig. 10 only three reconstructions without the characteristic residuals, as we did before. As can be seen there, the sampling method and the MAP method give very similar results in the region from 0.2 to 0.7 where the redshift space distortions are comparatively small, although the posterior mean tends to be slightly better, especially for $b = 1.5$. Similarly, the width of the error bars is virtually the same in this region.

This is not so in the region from 0.7 to 1.2. There, the MAP approach is taken in by the bifurcation and gives a very bad guess from 0.7 to 1.1 where it places peaks instead of valleys and a valley instead of a peak. What makes the situation problematic indeed is that the estimated error bars are small there. The algorithm therefore completely misjudges its accuracy. As already mentioned above, the situation improves slightly for the bias 1.5 but the problem is still there.

The sampling method, however, does the right thing in the interval 0.7–1.1. It nicely reshapens the largest peak and at the place of the side peak 0.75, which is in the ‘shadow’ of the larger peak at 0.9, it widens the error bar to remain on the safe side. We cannot judge for sure at this point whether this is just by chance unlucky data. For the larger bias 1.5, the improvement is not so good because the shadow on nearby structures is much stronger.

The impression that the reconstruction quality of the average map from Metropolis–Hastings sampling is much better than the $m_{cl}$ reconstruction also manifests itself in the $L_2$-distance, which we list in Table 3.

---

9 In order not to mix up the ranges $[a, b]$ and $[b, 1 \cup 0, a]$ we say $[b, 1 + a]$ if we mean the latter.
Note also how the uncertainty covariance matrix in Fig. 10 changes from $b = 0.5$ to $b = 1.5$. To understand what lies beneath, we have to look at the Gaussian approximation of the posterior $P(s|x) \approx \mathcal{N}(\mathcal{F}s-s^T\mathcal{D}^{-1}s/2)$. Here $\mathcal{N} = e^{-s^T/dj/2}/2\pi\mathcal{D}$ is a normalization factor and $j$ summarizes all terms of the probability Hamiltonian of first order in $s$. Direct calculation yields

$$
\langle s \rangle_{g(x)} \approx \mathcal{N} \int \mathcal{D}s \, e^{s^{T}j^{-1}s/2} = D \cdot j.
$$

Enßlin et al. (2009) call $j$ the information source because it excites the posterior mean from a zero-map to a non-zero state. We can show that in our problem $j$ contains a term linear in $d$ and additional terms. Therefore, equation (38) tells how the posterior mean reacts on additional data. Hence the pictures of the propagator in Fig. 10 show how the posterior map is constructed from the information source $j$ and introduces a correlation length in the map.

In void regions, the correlation length is longer than in overdense regions because the Poissonian noise is larger there. If there were no other source of uncertainty but the Poissonian noise, the band would be tighter in the most overdense regions. However, there is an additional source of uncertainty, namely the velocity dispersion from the redshift distortion model. Therefore, in overdense regions the correlation length of the propagator does not collapse, but shows a pattern of strong correlation and anticorrelation, as can be seen in the region from 0.7 to 1.2 in both cases. Comparing the correlation length to the distortion matrix in Fig. 9 shows that this is in good agreement with the correlation induced from the distortion matrix. The main difference in the correlation matrix of the example with $b = 1.5$ is that the contrast is sharper and that the diagonal is much more structured.

| $\bar{\rho}_{gal} = 500$ | $\bar{\rho}_{gal} = 1000$ | $\bar{\rho}_{gal} = 500$ |
|-------------------------|--------------------------|-------------------------|
| $b = 0.5$               | $b = 0.5$                | $b = 1.5$               |
| $\|m_{gal} - s_{\|}^{2}\|$ | 0.42                    | 0.41                    | 0.24                    |
| $\|s_{\|}^2 - \|s_{\|}^{2}\|$ | 0.17                    | 0.10                    | 0.16                    |

Table 3. $L_2$-distance of reconstruction to signal for three reconstructions, two of which are shown in Fig. 10. Note that there is hardly any improvement of the classical map from $\rho_{gal} = 1000$ to $\rho_{gal} = 1000$ for $b = 0.5$ while the posterior average improves by about 40 per cent. For the larger bias $b = 1.5$, the difference between $m_{gal}$ and the posterior average becomes less apparent.

4 CONCLUSIONS

Many reconstruction problems in cosmology suffer not only from large noise but also from substantial measurement uncertainties. While it is possible that some measurement uncertainties will be ameliorated in the future by more sophisticated techniques, other sources of uncertainty are fundamental, such as cosmic variance and galaxy redshift distributions. Areas where this applies are real space LSS reconstructions from galaxy counts in redshift space, but also consistent treatment of photometric redshift. For precision cosmology with galaxies, it is therefore of paramount importance to incorporate these uncertainties in the analysis.

Here we have presented a novel method for how spatially distorted lognormal fields as they occur in density field reconstruction can be reconstructed in a Bayesian way. This method was developed in the framework of IFT, which we outlined in Section 2. We have shown that the IFT moment calculation is based on the minimization of the expected $L_2$-weighted error of the reconstruction.

Where the exact moment calculation from the posterior was not possible, we argued how the correct map (the posterior mean) could be approximated by a MAP approach.

We developed a data model for a lognormal signal with Poissonian noise where the response can be non-local. We even allowed for the case in which the distortion of data space could depend on the signal that was to be reconstructed. The resulting problem is so complex that it could only be solved approximately via numerical minimization of its probability Hamiltonian.

For a test of our approach, we performed simulations where we constructed mock signals, produced mock data thereof and tried to reconstruct the underlying signal by numerical minimization of the probability Hamiltonian.

We tested our reconstruction code on three different distortion problems, as follows:

(i) data with typical distortions as they appear in photometric redshift measurement;
(ii) coded mask aperture problems as they appear in X-ray and $\gamma$-ray astronomy;
(iii) real space matter reconstruction from redshift distorted data.

For the latter, we developed a model for the forward problem to construct redshift space data from real space galaxy distributions and where the distortion was dependent on the underlying matter distribution that was to be measured. We were able to tackle this problem with a MAP approach. However, after further complication of the distortion operator we found that the MAP method does not live up to its expectations. Instead, we could show that approximating the posterior via Metropolis–Hastings sampling could give much more accurate reconstructions. Therefore, we think that for such complicated problems the MAP method gives misleading results and should be superseded by more powerful, but also computationally more demanding, approaches such as sampling the posterior PDF.

For the coded mask data, we were able to identify the largest peaks and we have shown that it is even possible to reconstruct their substructure if the count rates are high enough. An application of this approach to real X-ray or $\gamma$-ray data should be possible, but before doing so, some effort must be spent to make the approach more robust to the false detection of peaks.

At last, the reconstruction of a redshift space signal from photometric redshift data proved to be very fruitful. In many cases, we were able to reconstruct the underlying matter distribution remarkably well. As the colour space distortion is independent of the underlying signal, an application of our approach to large data sets is feasible.

We have also shown that in the IFT framework it is possible to easily combine data sets with different error characteristics. We considered the problem of combining photometric redshift data with large uncertainties and spectroscopic data that are very accurate in position. Our analysis has shown that even with a low abundance of accurate data it is possible to improve the reconstruction from distorted data with large abundance as long as there is room for improvement.

In all cases, we found that Bayesian analysis of the problem is inevitable for the noise level we were considering. We have also shown that the reconstruction becomes significantly worse when the data are distorted, but the data space distortion was neglected during the reconstruction. Therefore, we think that including the data space distortions in future precision analysis is inevitable.

Because the assumptions of our method are based on a few generic principles, we are confident that further areas will be found where our work will be appreciated.
ACKNOWLEDGMENTS

We thank Mona Frommert, Henrik Junklewitz, Jens Jasche, Niels Oppermann and an anonymous referee for discussions and comments on the manuscript.

REFERENCES

Benítez N., 2000, ApJ, 536, 571
Bertschinger E., Dekel A., Faber S. M., Dressler A., Burstein D., 1990, ApJ, 364, 370
Bialek W., Callan C. G., Strong S. P., 1996, Phys. Rev. Lett., 77, 4693
Binder K., 1997, Rep. Prog. Phys., 60, 487
Branchini E., Plionis M., Sciama D. W., 1996, ApJ, 461, L17
Bromley B. C., Warren M. S., Zurek W. H., 1997, ApJ, 475, 414
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Coles P., Jones B., 1991, MNRAS, 248, 1
Cornwell T. J., Evans K. F., 1985, A&A, 143, 77
Dicke R. H., 1968, ApJ, 153, L101
Duane S., Kennedy A. D., Pendleton B. J., Roweth D., 1987, Phys. Lett. B, 195, 216
Efstathiou G., 2000, MNRAS, 317, 697
Enßlin T., Frommert M., 2010, preprint (arXiv:1002.2928)
Erdogdu P. et al., 2004, MNRAS, 352, 939
Erdogdu P. et al., 2006, MNRAS, 373, 45
Frieden R. R., Wells D. C., 1978, J. Opt. Soc. Am. B, 68, 93
Fry J. N., Gaztanaga E., 1993, ApJ, 413, 447
Geman S., Geman D., 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721
Goldberg D. M., 2001, ApJ, 550, 87
Green P. et al., 2006, MNRAS, 373, 45
Hamilton A. S. J., 1995, in Maurogordato S., Balkowski C., Tao C., Tran ments on the manuscript.
Hamlet J. C., Uhlig J., Weiguny A., 2005, Eur. Phys. J. B, 46, 41
Heavens A. F., Taylor A. N., 1995, MNRAS, 275, 483
Hebert T. J., Leahy R., 1992, IEEE Trans. Signal Process., 40, 2290
Hastings W. K., 1970, Biometrika, 57, 97
Huchra J. et al., 2005, in Fairall A. P., Woudt P. A., eds, Proc. ASP Conf. Series Vol. 329, Nearby Large-Scale Structures and the Zone of Avoidance.
Huchra J. et al., 2005, in Fairall A. P., Woudt P. A., eds, Proc. ASP Conf. Series Vol. 329, Nearby Large-Scale Structures and the Zone of Avoidance.
Hamilton A. S. J., 1995, in Maurogordato S., Balkowski C., Tao C., Tran Thanh Van J., eds, Clustering in the Universe. Editions Frontières, Gif-sur-Yvette, p. 143
Hamilton A. S. J., 1998, in Hamilton D., ed., Astrophys. Space Sci. Library Vol. 231, The Evolving Universe. Kluwer, Dordrecht, p. 185
Hastings W. K., 1970, Biometrika, 57, 97
Heavens A. F., Taylor A. N., 1995, MNRAS, 275, 483
Hebert T. J., Leahy R., 1992, IEEE Trans. Signal Process., 40, 2290
Hoffman Y., Zaroubi S., 2000, ApJ, 535, L5
Hubble E., 1934, ApJ, 79, 8
Huchra J. et al., 2005, in Fairall A. P., Woudt P. A., eds, Proc. ASP Conf. Ser. Vol. 329, Nearby Large-Scale Structures and the Zone of Avoidance.
Astron. Soc. Pac., San Francisco, p. 135
Jasche J., Kitaura F. S., 2010, MNRAS, 407, 29
Jasche J., Kitaura F. S., Li C., Enßlin T. A., 2010, MNRAS, 409, 355
Jaynes E. T., 1957, Phys. Rev., 106, 620
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A.
Jasche J. et al., 2010, MNRAS, 403, 589
Landau L. D., Lifshitz E. M., 1966, Mechanik. Akademie-Verlag, Berlin
Layzer D., 1956, AJ, 61, 383
Lehm J. C., 1999, preprint (arXiv:physics/9912005)
Lemm J. C., 2001, in Mohammad-Djafari A., ed., AIP Conf. Ser. Vol. 568, Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Am. Inst. Phys., New York, p. 425
Lemm J. C., Uhlig J., Weiguny A., 2001, Eur. Phys. J. B, 20, 349
Lemm J. C., Uhlig J., Weiguny A., 2005, Eur. Phys. J. B, 46, 41
Martínez V. J., Saar E., 2002, Statistics of the Galaxy Distribution. Chapman & Hall, London
Matsubara T., 2008a, Phys. Rev. D, 78, 109901
Matsubara T., 2008b, Phys. Rev. D, 78, 083519
Mukhanov V., 2005, Physical Foundations of Cosmology. Cambridge Univ. Press, Cambridge
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Neyrinck M. C., Szapudi I., Szalay A. S., 2009, ApJ, 698, L90
Nishiyama R., Narayan R., 1982, JAA& 3, 419
Nunez J., Laclerc J., 1990, Ap&SS, 171, 341
Nusser A., Haehnelt M., 1999, MNRAS, 303, 179
Oh C., Frieden R. B., 2009, Opt. Commun., 282, 2489
Peacock J. A., Dodds S. J., 1994, MNRAS, 267, 1020
Peebles P. J. E., 1980, The Large-scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Percival W. J., 2005, MNRAS, 356, 1168
Percival W. J. et al., 2007, ApJ, 657, 645
Peskine M. E., Schroeder D. V., 1995, An Introduction to Quantum Field Theory. Westview Press, Boulder, CO
Ratcliffe A., Shanks T., Parker Q. A., Fong R., 1998, MNRAS, 296, 191
Schmidt D. I. M. et al., 1999, ApJ, 118, 1146
Shaya E. J., Peebles P. J. E., Tully R. B., 1995, ApJ, 454, 15
Shewchuk J. R., 1994, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. Available at http://www.cs.berkeley.edu/~quake-papers/painless-conjugate-gradient.ps
Smith R. E. et al., 2003, MNRAS, 341, 1311
Springel V. et al., 2005, Nat, 435, 629
Tadros H., Efstathiou G., 1996, MNRAS, 282, 1381
Tadros H. et al., 1999, MNRAS, 305, 527
Vogeley M. S., Hoyle F., Rojas R. R., Goldberg D. M., 2004, in Diaferio A., ed., Proc. IAU Colloq. 195, Outskirts of Galaxy Clusters: Intense Life in the Suburbs. Kluwer, Dordrecht, p. 5
Wang G., Fu L., Qi J., 2008, Phys. Medicine Biol., 53, 593
Webster M., Lahav O., Fisher K., 1997, MNRAS, 287, 425
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
Wittman D., 2009, ApJ, 700, L174
Yahi A., Strauss M. A., Davis M., Huchra J. P., 1991, ApJ, 372, 380

APPENDIX A: NOTATION

Having to deal extensively with field variables, we need a shorthand notation for the calculus. Hence, we define a vector space whose elements are fields; in this sense we use the terms ‘vector’ and ‘field’ interchangeably. The dot product in this vector space is defined by

\[ v^* w = \int dx v(x) w(x). \]  

(A1)

Instead of writing the field variable in round brackets we usually use a notation where the variable is in subscript to make the correspondence to finite dimensional vector spaces more obvious.

We also use tensors of higher rank over this vector space, most notably matrices. Analogously to equation (A1), we define

\[ (Mv)(x) = \int dy M(x,y)v(y) \]

\[ (MN)(x,y) = \int dz M(z,x)N(z,y), \]

and so on.

We also need a field space equivalent of the frequently used rules

\[ \frac{\partial}{\partial x_j} x_j = \delta_j, \quad \frac{\partial}{\partial x_i} \sum_j x_j k_j = k_i. \]

For all practical applications this is enough as we deal only with a finite number of pixels. However, all equations remain valid if we
define the functional derivative according to Peskin & Schroeder (1995, chapter 9) as
\[ \frac{\delta}{\delta J_k} J_j = \delta(x - y), \quad \frac{\delta}{\delta J_k} \int dy J_i \Phi_s = \Phi_s. \]
This blends in naturally with our definition of the vector product (A1).

Fields are printed in normal font with indices omitted. Matrices are in bold font, with the notable exception of derivatives of vectors such as $\delta\mu/\delta s$. To avoid an ambiguity in the sequence of the indices, we define for derivatives that the new index shall be last, i.e.
\[ \left( \frac{\delta \mu}{\delta s} \right)_{ij} = \frac{\delta \mu_j}{\delta s_i}. \]

All functions on fields are understood to act component-wise, that is $f(v)$ should be read as $f(v_i) \equiv f(v_i)$. In particular, this also applies to fundamental operations, such as multiplication and division: $(vw)_i \equiv v_i w_i$ and $(v/w)_i \equiv v_i / w_i$. Note also that we do not adopt Einstein’s summation convention, such that an expression such as $v e^{v_i}$ should in fact be read as the number $v_i$ multiplied by the number $e^{v_i}$.

In this notation, subtle differences do matter, but it shortens the formulas considerably. Here are some intricate examples:
\[ s^t S s = \sum_{i,j} S_{ij} s_i s_j \]
\[ [v(w)]_i = v_i \left( \sum_j S_{ij} w_j \right) \]
\[ (v w)_i = v_i w_i \]
\[ (v S)_i = v_i S_i \]
\[ (v^t S)_{ij} = \sum_j v_j S_{ji} \]

**APPENDIX B: GENERATING MOCK SIGNALS WITH THE GAUSSIAN COVARIANCE MATRIX**

The aim is to generate signals with the Gaussian covariance matrix
\[ \langle s^t S s \rangle_{(v)} = S. \]  
(B1)

Using this definition, $S$ is symmetric and positive definite, such that it has a root (i.e. there exists a matrix $T$ such that $S = T^t T$). We now show that the mock signal can be generated by convolving a vector of Gaussian random numbers $r$ with $T$: $s = Tr$.

**Reconstruction of spatially distorted fields**

Without loss of generality, we restrict our reasoning to the case where the Gaussian random numbers in $r$ have unit variance, such that the prior for $r$ is
\[ P(r) = \prod_i \frac{1}{\sqrt{2\pi}} e^{-r_i^2/2} = \frac{1}{\sqrt{2\pi |I|}} e^{-r^2/2}. \]  
(B2)

Here, the product runs over all pixel indices. From $r$ we construct the signal vector $s$ by application of $T$, which can be understood as a smoothing procedure for the random values. Note also that only in this step do different points of the signal become correlated with each other; the entries of $r$ are by virtue of the random nature of its entries uncorrelated.

We now have to prove that this procedure gives the desired prior for $s$:
\[ P(s) = \int Dr \delta(s - Tr) P(r) \]
\[ = \int Dq \int Dr \frac{1}{|2\pi|} e^{-q(s - T r)} e^{-r^2/2} \]
\[ = \frac{1}{|2\pi|} \int Dq e^{-q^t S^{-1} q - q^t T r} \]
\[ = \frac{1}{|2\pi S|^{1/2}} e^{-s^t S^{-1} s/2} = G(s, S), \]

As a source for random numbers we use the ‘Mersenne Twister’ random number generator, which offers a pseudo random number sequence with a period of $2^{19937} - 1$ (Matsumoto & Nishimura 1998). Its advantages are speed and very good randomness. In particular, we use its implementation from the GNU Scientific Library (gsl).\(^\text{10}\)

Note that this procedure gives an easy and robust way to test whether the constructed signal has indeed the desired covariance $s^t S^{-1} s = r^t T^t S^{-1} T r = r^t r \approx N_{\text{pix}}$, where we have used $r$ as a vector of $N_{\text{pix}}$ Gaussian random numbers of unit variance in the last step. In other words, for a Gaussian random field $s$, the number $s^t S^{-1} s$ should roughly give the dimension of $s$.

\(^{10}\)The GNU Scientific Library is available from http://www.gnu.org/software/gsl

This paper has been typeset from a T\(\hat{\text{e}}\)X/L\(\hat{\text{e}}\)X file prepared by the author.