Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **407,** 29–42 (2010) doi:10.1111/j.1365-2966.2010.16897.x

Fast Hamiltonian sampling for large-scale structure inference

Jens Jasche[1]★ and Francisco S. Kitaura[2]

[1]*Max-Planck-Institut für Astrophysik , Karl-Schwarzschild Straße 1, D-85748 Garching, Germany*
[2]*SISSA, Scuola Internazionale Superiore di Studi Avanzati, via Beirut 2-4, I-34014 Trieste, Italy*

Accepted 2010 April 21. Received 2010 March 19; in original form 2009 September 21

ABSTRACT

In this paper we present a new and efficient Bayesian method for non-linear three-dimensional large-scale structure inference. We employ a Hamiltonian Monte Carlo (HMC) sampler to obtain samples from a multivariate highly non-Gaussian lognormal Poissonian density posterior given a set of observations. The HMC allows us to take into account the non-linear relations between the observations and the underlying density field which we seek to recover. As the HMC provides a sampled representation of the density posterior any desired statistical summary, such as the mean, mode or variance, can be calculated from the set of samples. Further, it permits us to seamlessly propagate non-Gaussian uncertainty information to any final quantity inferred from the set of samples. The developed method is extensively tested in a variety of test scenarios, taking into account a highly structured survey geometry and selection effects. Tests with a mock galaxy catalogue based on the Millennium Run show that the method is able to recover the filamentary structure of the non-linear density field. The results further demonstrate the feasibility of non-Gaussian sampling in high-dimensional spaces, as required for precision non-linear large-scale structure inference. The HMC is a flexible and efficient method, which permits for simple extension and incorporation of additional observational constraints. Thus, the method presented here provides an efficient and flexible basis for future high-precision large-scale structure inference.

Key words: methods: data analysis – galaxies: statistics – cosmology: observations – large scale structure of Universe.

1 INTRODUCTION

Modern large galaxy surveys allow us to probe cosmic large-scale structure to very high accuracy if the enormous amount of data can be processed and analysed efficiently. Especially, precision reconstruction of the three-dimensional density field from observations poses complex numerical challenges. For this reason, several reconstruction methods and attempts to recover the underlying density field from galaxy observations have been presented in literature (see e.g. Bertschinger & Dekel 1989, 1991; Hoffman 1994; Lahav et al. 1994; Fisher et al. 1995; Sheth 1995; Webster, Lahav & Fisher 1997; Bistolas & Hoffman 1998; Schmoldt et al. 1999; Saunders & Ballinger 2000a; Zaroubi 2002; Erdoğdu et al. 2004; Kitaura et al. 2009; Jasche et al. 2010). Recently, Kitaura et al. (2009) presented a high-resolution Wiener reconstruction of the Sloan Digital Sky Survey (SDSS) matter density field, and demonstrated the feasibility of precision large-scale structure analysis. The Wiener filtering approach is based on a linear data model, which takes into account several observational effects, such as survey geometry, selection effects and noise (Kitaura & Enßlin 2008; Kitaura et al. 2009; Jasche et al. 2010). Although, the Wiener filter has proven to be extremely efficient for three-dimensional matter field reconstruction, it still relies on a Gaussian approximation of the density posterior. While this is an adequate approximation for the largest scales, precision recovery of non-linear density structures may require non-Gaussian posteriors. Especially, the detailed treatment of the non-Gaussian behaviour and structure of the Poissonian shot-noise contribution may allow for more precise recovery of poorly sampled objects. In addition, for a long time it has been suggested that the fully evolved non-linear matter field can be well described by lognormal statistics (see e.g. Hubble 1934; Peebles 1980; Coles & Jones 1991; Gaztanaga & Yokoyama 1993; Kayo, Taruya & Suto 2001). These discussions seem to advocate the use of a lognormal Poissonian posterior for large-scale structure inference. Several methods have been proposed to take into account non-Gaussian density posteriors (see e.g. Saunders & Ballinger 2000b; Kitaura & Enßlin 2008; Enßlin, Frommert & Kitaura 2009).

However, if the recovered non-linear density field is to be used for scientific purposes, the method not only has to provide a single estimate, such as a mean or maximum a posteriori reconstruction, but it should also provide uncertainty information, and the means to

★E-mail: jjasche@mpa-garching.mpg.de

© 2010 The Authors. Journal compilation © 2010 RAS

non-linearly propagate this uncertainty to any final quantity inferred from the recovered density field.

For this reason, here we propose a new Bayesian method for non-linear large-scale structure inference. The developed computer program HAmiltonian Density Estimation and Sampling (HADES) explores the posterior distribution via a Hamiltonian Monte Carlo (HMC) sampling scheme. Unlike conventional Metropolis–Hastings algorithms, which move through the parameter space by a random walk, and therefore require prohibitive amounts of steps to explore high-dimensional spaces, the HMC sampler suppresses random walk behaviour by introducing a persistent motion of the Markov chain through the parameter space (Duane et al. 1987; Neal 1993, 1996). In this fashion, the HMC sampler maintains a reasonable efficiency even for high-dimensional problems (Hanson 2001). The HMC sampler has been widely used in Bayesian computation (see e.g. Neal 1993). In cosmology it has been employed for cosmological parameter estimation and cosmic microwave background data analysis (Hajian 2007; Taylor, Ashdown & Hobson 2008).

In this paper we demonstrate that the HMC can efficiently be used to sample the lognormal Poissonian posterior even in high-dimensional spaces. In this fashion, the method is able to take into account the non-linear relationship between the observation and the underlying density which we seek to recover. The scientific output of the HMC is a sampled representation of the density posterior. For this reason, any desired statistical summary such as mean, mode and variance can easily be calculated from the HMC samples. Further, the full non-Gaussian uncertainty information can seamlessly be propagated to any finally estimated quantity by simply applying the corresponding estimation procedure to all samples. This allows us to estimate the accuracy of conclusions drawn from the analysed data.

In this paper, we begin, in Section 2, by presenting a short justification for the use of the lognormal distribution as a prior for non-linear density inference, followed by a discussion of the lognormal Poissonian posterior in Section 3. Section 4 outlines the HMC method. In Section 5 the Hamiltonian equations of motion for the lognormal Poissonian posterior are presented. Details of the numerical implementation are described in Section 6. The method is extensively tested in Section 7 by applying HADES to generated mock observations, taking into account a highly structured survey geometry and selection effects. In Section 8 we summarize and conclude.

2 THE LOGNORMAL DISTRIBUTION OF DENSITY

In standard cosmological pictures, it is assumed that the initial seed perturbations in the primordial density field originated from an inflationary phase in the early stages of the big bang. This inflationary phase enhances microscopic quantum fluctuations to macroscopic scales yielding the initial density fluctuations required for gravitational collapse. These theories predict the initial density field amplitudes to be Gaussian distributed. However, it is obvious that Gaussianity of the density field can only be true in the limit $|\delta| \ll 1$, where δ is the density contrast. In fully evolved density fields with amplitudes of $\sigma_8 > 1$, as observed in the sky at scales of galaxies, clusters and superclusters, a Gaussian density distribution would allow for negative densities, and therefore would violate weak and strong energy conditions. In particular, it would give rise to negative mass ($\delta < -1$). Therefore, in the course of gravitational structure formation the density field must have changed its statistical properties. Coles & Jones (1991) argue that assuming Gaussian initial conditions in the density and velocity distributions will lead to a lognormally distributed density field. It is a direct consequence of the continuity equation or the conservation of mass.

Although, the exact probability distribution for the density field in non-linear regimes is not known, the lognormal distribution seems to be a good phenomenological guess with a long history. Already Hubble noted that galaxy counts in two-dimensional cells on the sky can be well approximated by a lognormal distribution (Hubble 1934). Subsequently, the lognormal distribution has been extensively discussed and agreements with galaxy observations have been found (e.g. Hubble 1934; Peebles 1980; Coles & Jones 1991; Gaztanaga & Yokoyama 1993; Kayo et al. 2001). Kayo et al. (2001) studied the probability distribution of cosmological non-linear density fluctuations from N-body simulations with Gaussian initial conditions. They found that the lognormal distribution accurately describes the non-linear density field even up to values of the density contrast of $\delta \sim 100$.

Therefore, according to observations and theoretical considerations, we believe that the statistical behaviour of the non-linear density field can be well described by a multivariate lognormal distribution, as given by

$$\mathcal{P}(\{s_k\}|\mathbf{Q}) = \frac{1}{\sqrt{2\pi \det(\mathbf{Q})}}\, \mathrm{e}^{-(1/2)\sum_{ij}[\ln(1+s_i)+\mu_i]\mathbf{Q}^{-1}_{ij}[\ln(1+s_j)+\mu_j]} \times \prod_k \frac{1}{1+s_k}, \tag{1}$$

where $\mathbf{Q}$ is the covariance matrix of the lognormal distribution and μ_i describes a constant mean field given by

$$\mu_i = \frac{1}{2}\sum_{i,j} \mathbf{Q}_{ij}. \tag{2}$$

This probability distribution seems to be an adequate prior choice for reconstructing the present density field. However, using such a prior requires highly non-linear reconstruction methods, as will be presented in the following.

3 LOGNORMAL POISSONIAN POSTERIOR

Studying the actual matter distribution of the Universe requires to draw inference from some observable tracer particle. The most obvious tracer particles for the cosmic density field are galaxies, which tend to follow the gravitational potential of matter. As galaxies are discrete particles, the galaxy distribution can be described as a specific realization drawn from an inhomogeneous Poisson process (see e.g. Layzer 1956; Peebles 1980; Martínez & Saar 2002). The corresponding probability distribution is given as

$$\mathcal{P}\left(\left\{N_k^{\mathrm{g}}\right\}|\{\lambda_k\}\right) = \prod_k \frac{(\lambda_k)^{N_k^{\mathrm{g}}}\, \mathrm{e}^{-\lambda_k}}{N_k^{\mathrm{g}}!}, \tag{3}$$

where N_k^{g} is the observed galaxy number at position $\boldsymbol{x}_k$ in the sky and λ_k is the expected number of galaxies at this position. The mean galaxy number is related to the signal s_k via

$$\lambda_k = R_k \bar{N}[1 + B(s)_k], \tag{4}$$

where R_k is a linear response operator, incorporating survey geometries and selection effects, $\bar{N}$ is the mean number of galaxies in the volume and $B(x)_k$ is a non-linear, non-local, bias operator at position $\boldsymbol{x}_k$. The lognormal prior given in equation (1) together with the Poissonian likelihood given in equation (3) yields the lognormal Poissonian posterior, for the density contrast s_k given some galaxy

observations $N_k^{\rm g}$:

$$\mathcal{P}\left(\{s_k\}|\left\{N_k^{\rm g}\right\}\right) = \frac{{\rm e}^{-(1/2)\sum_{ij}[\ln(1+s_i)+\mu_i]\mathbf{Q}_{ij}^{-1}[\ln(1+s_j)+\mu_j]}}{\sqrt{2\pi\det(\mathbf{Q})}}\prod_l \frac{1}{1+s_l} \times \prod_k \frac{[R_k\bar{N}(1+B(s)_k)]^{N_k^{\rm g}}\,{\rm e}^{-R_k\bar{N}[1+B(s)_k]}}{N_k^{\rm g}!} \quad (5)$$

However, this posterior greatly simplifies if we perform the change of variables by introducing $r_k = \ln(1+s_k)$. Note that this change of variables is also numerically advantageous, as it prevents numerical instabilities at values $\delta \sim -1$. Hence, we yield the posterior

$$\mathcal{P}\left(\{r_k\}|\left\{N_k^{\rm g}\right\}\right) = \frac{{\rm e}^{-(1/2)\sum_{ij}(r_i+\mu_i)\mathbf{Q}_{ij}^{-1}(r_j+\mu_j)}}{\sqrt{2\pi\det(\mathbf{Q})}} \times \prod_k \frac{\left\{R_k\bar{N}\left[1+B({\rm e}^r-1)_k\right]\right\}^{N_k^{\rm g}}\,{\rm e}^{-R_k\bar{N}[1+B({\rm e}^r-1)_k]}}{N_k^{\rm g}!}. \quad (6)$$

It is important to note that this is a highly non-Gaussian distribution, and non-linear reconstruction methods are required in order to perform accurate matter field reconstructions in the non-linear regime. In example, estimating the maximum a posteriori values from the lognormal Poissonian distribution involves the solution of implicit equations. However, we are not solely interested in a single estimate of the density distribution; we rather prefer to draw samples from the lognormal Poissonian posterior. In the following, we are therefore describing a numerically efficient method to sample from this highly non-Gaussian distribution.

4 HAMILTONIAN SAMPLING

As already described in the previous section the lognormal Poissonian posterior will involve highly non-linear reconstruction methods and will therefore be numerically demanding. Nevertheless, since we propose a Bayesian method, we are not interested in only providing a single estimate of the density field, but would rather be able to sample from the full non-Gaussian posterior. Unlike, in the Gibbs sampling approach to density field sampling, as proposed in Jasche et al. (2010), there unfortunately exists no known way to directly draw samples from the lognormal Poissonian distribution. For this reason, a Metropolis–Hastings sampling mechanism has to be employed.

However, the Metropolis–Hastings has the numerical disadvantage that not every sample will be accepted. A low acceptance rate can therefore result in a prohibitive numerical scaling for the method, especially since we are interested in estimating full three-dimensional matter fields which usually have about 10^6 or more free parameters s_k. This high rejection rate is due to the fact that conventional Markov chain Monte Carlo (MCMC) methods move through the parameter space by a random walk and therefore require a prohibitive amount of samples to explore high-dimensional spaces. Given this situation, we propose to use a hybrid Monte Carlo method which, in the absence of numerical errors, would yield an acceptance rate of unity (see e.g. Duane et al. 1987; Neal 1993, 1996).

The so-called HMC method exploits techniques developed to follow classical dynamical particle motion in potentials (Duane et al. 1987; Neal 1993, 1996). In this fashion the Markov sampler follows a persistent motion through the parameter space, suppressing the random walk behaviour. This enables us to sample with reasonable efficiency in high-dimensional spaces (Hanson 2001).

The idea of the Hamiltonian sampling can be easily explained. Suppose that we wish to draw samples from the probability distribution $\mathcal{P}(\{x_i\})$, where $\{x_i\}$ is a set consisting of the N elements x_i. If we interpret the negative logarithm of this posterior distribution as a potential,

$$\psi(x) = -\ln(\mathcal{P}(x)), \quad (7)$$

and by introducing a 'momentum' variable p_i and a 'mass matrix' $\mathbf{M}$, as nuisance parameters, we can formulate a Hamiltonian describing the dynamics in the multidimensional phase space. Such a Hamiltonian is then given as

$$H = \sum_i\sum_j \frac{1}{2}\,p_i\,\mathbf{M}_{ij}^{-1}\,p_j + \psi(x). \quad (8)$$

As can be seen in equation (8), the form of the Hamiltonian is such that this distribution is separable into a Gaussian distribution in the momenta $\{p_i\}$ and the target distribution $\mathcal{P}(\{x_i\})$ as

$${\rm e}^{-H} = \mathcal{P}(\{x_i\})\,{\rm e}^{-(1/2)\sum_i\sum_j\,p_i\,\mathbf{M}_{ij}^{-1}\,p_j}. \quad (9)$$

It is therefore obvious that marginalizing over all momenta will yield again our original target distribution $\mathcal{P}(\{x_i\})$.

Our task now is to draw samples from the joint distribution, which is proportional to $\exp(-H)$. To find a new sample of the joint distribution we first draw a set of momenta from the distribution defined by the kinetic energy term, which is an N-dimensional Gaussian with a covariance matrix $\mathbf{M}$. We then allow our system to evolve deterministically, from our starting point $(\{x_i\}, \{p_i\})$ in the phase space for some fixed pseudo-time τ according to Hamilton's equations:

$$\frac{{\rm d}x_i}{{\rm d}t} = \frac{\partial H}{\partial p_i}, \quad (10)$$

$$\frac{{\rm d}p_i}{{\rm d}t} = \frac{\partial H}{\partial x_i} = -\frac{\partial\psi(x)}{\partial x_i}. \quad (11)$$

The integration of these equations of motion yields the new position $(\{x_i'\}, \{p_i'\})$ in phase space. This new point is accepted according to the usual acceptance rule

$$\mathcal{P}_A = \min[1, \exp(-(H(\{x_i'\}, \{p_i'\}) - H(\{x_i\}, \{p_i\}))]. \quad (12)$$

Since the equations of motion provide a solution to a Hamiltonian system, energy or the Hamiltonian given in equation (8) is conserved, and therefore the solution to this system provides an acceptance rate of unity. In practice however, numerical errors can lead to a somewhat lower acceptance rate. Once a new sample has been accepted the momentum variable is discarded and the process restarts by randomly drawing a new set of momenta. The individual momenta $\{p_i\}$ will not be stored, and therefore discarding them amounts to marginalizing over this auxiliary quantity. Hence, the Hamiltonian sampling procedure basically consists of two steps. The first step is a Gibbs sampling step, which yields a new set of Gaussian-distributed momenta. The second step, on the other hand, amounts to solving a dynamical trajectory on the posterior surface. In this fashion, the HMC incorporates a particular Metropolis–Hastings transition probability which, in the absence of numerical errors, guarantees an acceptance rate of unity (Neal 1993, 1996). However, in a numerical implementation rejection may occur in cases where the numerical integration of the Hamiltonian trajectory is not accurate enough. In this case, the Metropolis acceptance rule given in equation (12) will restore detailed balance and guarantees the chain to converge to the true target distribution $\mathcal{P}(\{x_i\})$.

The efficiency of the HMC is thus partly determined by the accuracy of the numerical integration scheme. The particular numerical implementation of our method will be discussed below.

5 EQUATIONS OF MOTION FOR A LOGNORMAL POISSONIAN SYSTEM

In the framework of Hamiltonian sampling the task of sampling from the lognormal Poissonian posterior reduces to solving the corresponding Hamiltonian system. Given the posterior distribution defined in equation (6) we can write the potential $\psi(\{r_k\})$ as

$$\begin{aligned}\psi(\{r_k\}) = &\frac{1}{2}\ln(2\pi\det(\mathbf{Q})) \\ &+\frac{1}{2}\sum_{ij}(r_i+\mu_i)\mathbf{Q}^{-1}_{ij}(r_j+\mu_j) \\ &-\sum_k\left\{\ln\left(\frac{(R_k\bar{N})^{N^{\rm g}_k}}{N^{\rm g}_k!}\right)+N^{\rm g}_k\,\ln(1+B(\mathrm{e}^r-1)_k)\right. \\ &\left.-R_k\bar{N}\left[1+B(\mathrm{e}^r-1)_k\right]\right\}. \end{aligned} \tag{13}$$

The gradient of this potential with respect to r_l then yields the forces, given as

$$\begin{aligned}\frac{\partial\psi(\{r_k\})}{\partial r_l} = &\sum_j \mathbf{Q}^{-1}_{lj}(r_j+\mu_j) \\ &-\left(\frac{N^{\rm g}_l}{1+B(\mathrm{e}^r-1)_l}-R_l\bar{N}\right)\left.\frac{\partial B(\mathrm{e}^r-1)}{\partial(\mathrm{e}^r-1)}\right|_l \mathrm{e}^{r_l}. \end{aligned} \tag{14}$$

Equation (14) obviously is a very general formulation of the reconstruction problem, and it demonstrates that the Hamiltonian sampler can in principle deal with all kinds of non-linearities, especially in the case of the bias operator $B(x)$. However, for the sake of this paper, but without loss of generality, in the following we will assume a linear bias model, as a special case.

There are several things to remark about the treatment of galaxy bias at this point. Most importantly note that our method does not require the definition of a linear bias. In order to take into account more general non-linear and non-local bias models, one simply solves the problem according to equations (13) and (14) by specifying the desired bias model and its derivative. However, the sampling method, as described in this paper, provides a much more appealing approach to study various different bias models via a Blackwell–Rao estimator. As described in Appendix C this procedure allows for the generation of various density posterior distributions with different bias models in a post-processing step. Since the generation of these Blackwell–Rao estimators requires only a negligible fraction of the total computation time required to produce the Markov chain, it is much more appealing to defer the treatment of various different bias models to the post-processing step, rather than running many Markov chains each with a different bias model to study, although this can be done for the expense of additional computational time. Since the bias treatment can in principle be deferred to a post-processing analysis and since the linear bias model is computationally less expensive than a more complex model in the sampling procedure, in the following we will assume the linear bias relation

$$B(x)_k = bx_k, \tag{15}$$

where b is a constant linear bias factor.

We then obtain the potential

$$\begin{aligned}\psi(\{r_k\}) = &\frac{1}{2}\ln(2\pi\det(\mathbf{Q})) \\ &+\frac{1}{2}\sum_{ij}(r_i+\mu_i)\mathbf{Q}^{-1}_{ij}(r_j+\mu_j) \\ &-\sum_k\left\{\ln\left(\frac{(R_k\bar{N})^{N^{\rm g}_k}}{N^{\rm g}_k!}\right)+N^{\rm g}_k\,\ln(1+b\,(\mathrm{e}^r-1))\right. \\ &\left.-R_k\bar{N}[1+b\,(\mathrm{e}^r-1)]\right\}, \end{aligned} \tag{16}$$

and the corresponding gradient reads

$$\frac{\partial\psi(\{r_k\})}{\partial r_l} = \sum_j \mathbf{Q}^{-1}_{lj}(r_j+\mu_j)-\left(\frac{N^{\rm g}_l}{1+b\,(\mathrm{e}^r-1)}-R_l\bar{N}\right)b\,\mathrm{e}^{r_l}. \tag{17}$$

Inserting these results in equations (10) and (11) then yields the equations of motion

$$\frac{\mathrm{d}r_i}{\mathrm{d}t} = \sum_j \mathbf{M}^{-1}_{ij}\,p_j \tag{18}$$

and

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\sum_j \mathbf{Q}^{-1}_{ij}(r_j+\mu_j)-\left(\frac{N^{\rm g}_i}{1+b\,(\mathrm{e}^{r_i}-1)}-R_i\bar{N}\right)b\,\mathrm{e}^{r_i}. \tag{19}$$

New points on the lognormal Poissonian posterior surface can then easily be obtained by solving for the trajectory governed by the dynamical equations (18) and (19).

6 NUMERICAL IMPLEMENTATION

Our numerical implementation of the lognormal Poissonian sampler is named HADES. It utilizes the FFTW3 library for fast Fourier transforms and the GNU scientific library (gsl) for random number generation (Galassi et al. 2003; Frigo & Johnson 2005). In particular, we use the Mersenne Twister MT19937, with 32-bit word length, as provided by the gsl_rng_mt19937 routine, which was designed for Monte Carlo simulations (Matsumoto & Nishimura 1998).

6.1 The leapfrog scheme

As described above, a new sample can be obtained by calculating a point on the trajectory governed by equations (18) and (19). This means that if we are able to integrate the Hamiltonian system exactly energy will be conserved along such a trajectory, yielding a high probability of acceptance. However, the method is more general due to the Metropolis acceptance criterion given in equation (12). In fact, it is allowed to follow any trajectory to generate a new sample. This would enable us to use approximate Hamiltonians, which may be evaluated computationally more efficiently. Note, however, that only trajectories that approximately conserve the Hamiltonian given in equation (8) will result in high acceptance rates.

In order to achieve an optimal acceptance rate, we seek to solve the equations of motion exactly. For this reason, we employ a leapfrog scheme for the numerical integration. Since the leapfrog is a symplectic integrator, it is exactly reversible, a property required to ensure that the chain satisfies detailed balance (Duane et al. 1987). It is also numerically robust, and allows for simple propagation of errors. Here, we will implement the kick–drift–kick scheme. The

equations of motions are integrated by making n steps with a finite step size ϵ, such that $\tau = n\epsilon$:

$$p_i\left(t+\frac{\epsilon}{2}\right) = p_i(t) - \frac{\epsilon}{2}\left.\frac{\partial\psi(\{r_k\})}{\partial r_l}\right|_{r_i(t)}, \tag{20}$$

$$r_i\,(t+\epsilon) = r_i(t) - \frac{\epsilon}{m_i}\,p_i\left(t+\frac{\epsilon}{2}\right), \tag{21}$$

$$p_i\,(t+\epsilon) = p_i\left(t+\frac{\epsilon}{2}\right) - \frac{\epsilon}{2}\left.\frac{\partial\psi(\{r_k\})}{\partial r_l}\right|_{r_i(t+\epsilon)}. \tag{22}$$

We iterate these equations until $t = \tau$. Also note that it is important to vary the pseudo-time interval τ to avoid resonant trajectories. We do so by drawing n and ϵ randomly from a uniform distribution. For the time being we will employ the simple leapfrog scheme. However, it is possible to use higher order integration schemes, provided that exact reversibility is maintained.

6.2 Hamiltonian mass

The Hamiltonian sampler has a large number of adjustable parameters, namely the Hamiltonian 'mass matrix, $\mathbf{M}$, which can greatly influence the sampling efficiency. If the individual r_k were Gaussian distributed, a good choice for HMC masses would be to set them inversely proportional to the variance of that specific r_k (Taylor et al. 2008). However, for non-Gaussian distributions, such as the lognormal Poissonian posterior, it is reasonable to use some measure of the width of the distribution (Taylor et al. 2008). Neal (1996) proposes to use the curvature at the peak.

In our case, we expanded the Hamiltonian given in equation (16) in a Taylor series up to quadratic order for $|r_i| \ll 1$. This Taylor expansion yields a Gaussian approximation of the lognormal Poissonian posterior. Given this approximation and according to the discussion in Appendix A, the Hamiltonian mass should be set as

$$\mathbf{M}_{ij} = \mathbf{Q}_{ij}^{-1} - \left[\left(N_i^{\mathrm{g}} - R_i\bar{N}\right)b - N_i^{\mathrm{g}}b^2\right]\delta_{ij}^K. \tag{23}$$

However, calculation of the leapfrog scheme requires inversions of $\mathbf{M}$. Considering the high dimensionality of the problem, inverting and storing $\mathbf{M}^{-1}$ is computationally impractical. For this reason we construct a diagonal 'mass matrix' from equation (23). We found that choosing the diagonal of $\mathbf{M}$, as given in equation (23), in its Fourier basis yields faster convergence for the sampler than a real-space representation since it accounts for the correlation structure of the underlying density field.

6.3 Parallelization

For any three-dimensional sampling method, such as the lognormal Poisson sampler or the Gibbs sampler presented in Jasche et al. (2010), CPU time is the main limiting factor. For this reason parallelization of the code is a crucial issue. Since our method is a true Monte Carlo method, there exist in principle two different approaches to parallelize our code.

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

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

7 TESTING HADES

In this section, we apply HADES to simulated mock observations, where the underlying matter signal is perfectly known. With these tests we will be able to demonstrate that the code produces results consistent with the theoretical expectation. Furthermore, we wish to gain insight into how the code performs in real-world applications, when CPU time is limited.

7.1 Setting up mock observations

In this section we will describe a similar testing set-up as described in Jasche et al. (2010). For the purpose of this paper we generate lognormal random fields according to the probability distribution given in equation (1). These lognormal fields are generated based on cosmological power spectra for the density contrast δ. We generate these power spectra, with baryonic wiggles, following the prescription described in Eisenstein & Hu (1998) and Eisenstein & Hu (1999) and assuming a standard Lambda cold dark matter (ΛCDM) cosmology with the set of cosmological parameters ($\Omega_{\mathrm{m}} = 0.24, \Omega_\Lambda = 0.76, \Omega_{\mathrm{b}} = 0.04, h = 0.73, \sigma_8 = 0.74, n_s = 1$). Given these generated density fields we draw random Poissonian samples according to the Poissonian process described in equation (3).

The survey properties are described by the galaxy selection function F_i and the observation mask $\mathbf{M}_i$ where the product

$$R_i = F_i\,\mathbf{M}_i \tag{24}$$

yields the linear response operator.

The selection function is given by

$$F_i = \left(\frac{r_i}{r_0}\right)^b \left(\frac{b}{\gamma}\right)^{-b/\gamma} \mathrm{e}^{b/\gamma-(r_i/r_0)^\gamma}, \tag{25}$$

where r_i is the comoving distance from the observer to the centre of the ith voxel. For our simulation we chose parameters $b = 0.6$, $r_0 = 500\,\mathrm{Mpc}$ and $\gamma = 2$.

In Fig. 1 we show the selection function together with the sky mask, which defines the observed regions in the sky. The two-dimensional sky mask is given in sky coordinates of right ascension and declination. We designed the observational mask to represent some of the prominent features of the SDSS mask (see Abazajian et al. 2009, for a description of the SDSS Data Release 7). The projection of this mask into the three-dimensional volume yields the three-dimensional mask $\mathbf{M}_i$.

Two different projections of this generated mock galaxy survey are presented in Fig. 2 to give a visual impression of the artificial galaxy observation.

7.2 Burn-in behaviour

The theory described above demonstrates that the Hamiltonian sampler will provide samples from the correct probability distribution function as long as the initial conditions for the leapfrog integration are part of the posterior surface. However, in practice the sampler is not initialized with a point on the posterior surface, and therefore an initial burn-in phase is required until a point on the correct

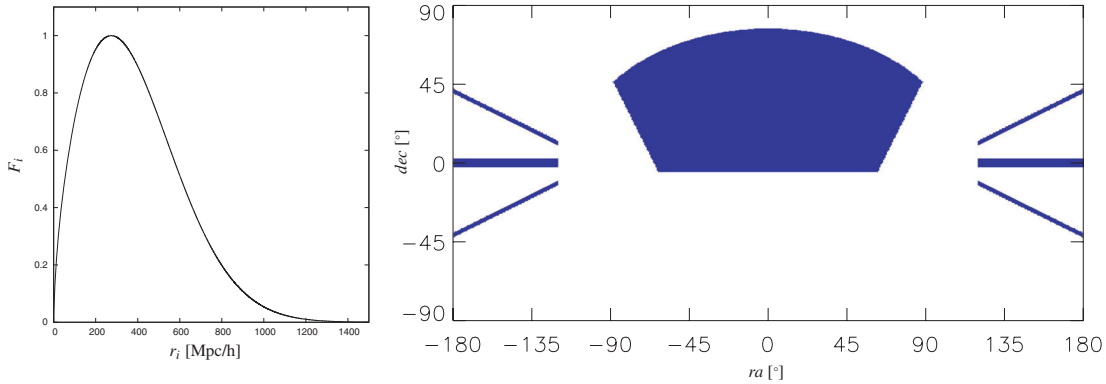

Figure 1. Selection function and two-dimensional sky mask used for the generation of mock galaxy observations.

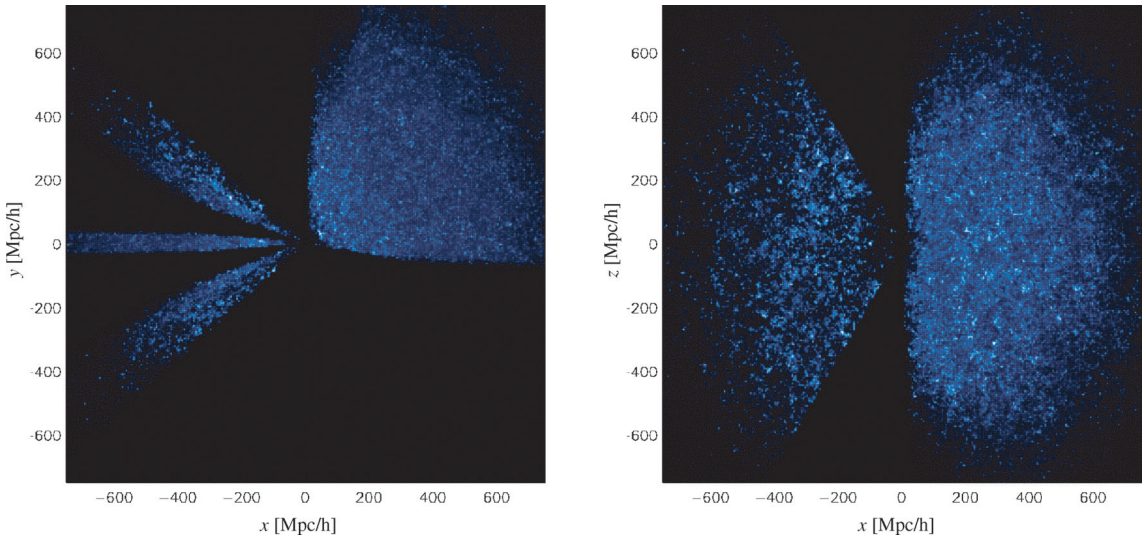

Figure 2. Volume rendering of artificial galaxy counts, generated as described in Section 7.1. The two panels show different projections. Effects of survey geometry and selection function are clearly visible. The observer is centred at (0, 0, 0).

posterior surface is identified. As there exists no theoretical criterion, which tells us when the initial burn-in period is completed, we have to test this initial sampling phase through experiments. These experiments are of practical relevance for real-world applications, as they allow us to estimate how many samples are required before the sampler starts sampling from the correct posterior distribution. To gain intuition we set up a simple experiment, in which we set the initial guess for the lognormal field constant to unity ($r_k^0 = 1$). Therefore, the initial samples in the chain will be required to recover structures contained in the observation. In order to gain intuition for the behaviour of our non-linear Hamiltonian sampler, we compare two cases. The first case consists of an artificial observation including selection effects and observational mask generated as described above. The second case is a comparison calculation, where we set the observation response operator $R_i = 1$. In this latter fiducial case, only shot noise remains as observational uncertainty. It is important to note that the individual Markov samples are unbiased in the sense that they possess the correct power information. Unlike a filter, which suppresses power in the low signal-to-noise ratio regions, the Hamiltonian sampler draws true samples from the lognormal Poissonian posterior, given in equation (5), once burn-in is completed. Therefore, a lognormal Poissonian sample has to be understood as consisting of a true signal part, which can be extracted from the observation and a fluctuating component, which restores power lost due to the observation. This interpretation is similar to the behaviour of the Gibbs sampler, as discussed in Jasche et al. (2010), with the exception that there is no obvious way to separate the true signal part from the variance contribution for the non-linear Hamiltonian sampler. Hence, the lower the signal-to-noise ratio of the data, the higher will be the fluctuating component.

This effect can be observed in Fig. 3 where we compare three successive Markov samples to the true mock signal via a point-to-point statistics. It can be nicely seen that the correlation with the true underlying mock signal improves as burn-in progresses. As expected, the fiducial calculation, shown in the right-hand panels of Fig. 3, has a much better correlation with the underlying true mock signal than the full observation. This is clearly owing to the fact that the full observation introduces much more variance than in the fiducial case. To visualize this fact further, we calculate the Euclidean distance between Markov samples and the true mock signal,

$$d_k\left(\left\{\delta^i_{\rm true}\right\},\left\{\delta^i_k\right\}\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\delta^i_{\rm true}-\delta^i_k\right)^2}, \tag{26}$$

over the progression of the Markov chain. In the lower panels of Fig. 3, it can be observed that the Eucledian distance drops initially and then saturates at a constant minimal d_k. This minimal d_k is

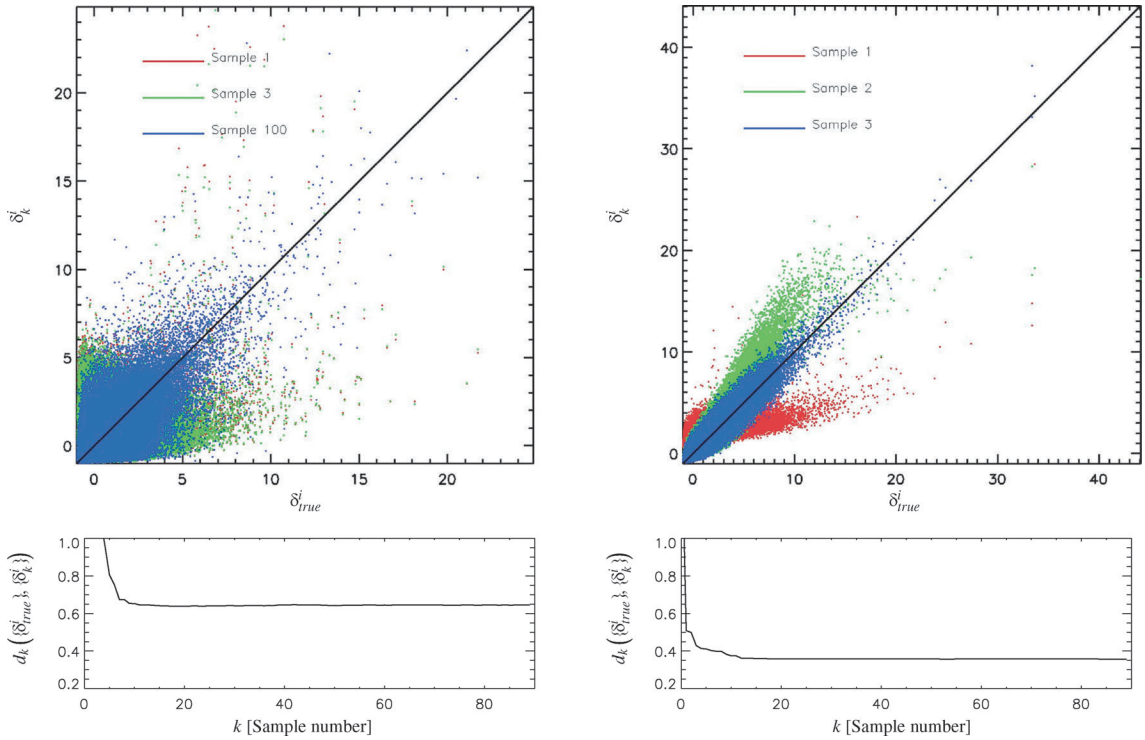

Figure 3. Tests of the initial burn-in behaviour for the two test cases of the fiducial calculation (right-hand panels) and the full test taking into account the observational uncertainties of survey geometry and selection effects (left-hand panels). The upper panels show successive point-to-point statistics between the individual samples and the true underlying mock signal. It can be seen that the successive Hamiltonian samples show increasing correlation with the true underlying signal. The lower panels show the successive Euclidean distances between samples and the true underlying signal during burn-in.

related to the intrinsic variance contribution in the individual samples. While the variance is lower for the fiducial observation, it is higher for the full observation.

As HADES produces unbiased samples, we can gain more detailed insight into the initial burn-in phase of the Markov chain, by following the evolution of successive power spectra measured from the samples. In addition, we measure the deviation ξ_l^k of the sample power spectra P_l^k to the power spectrum of the true mock matter field realization $P_l^{\rm true}$ via

$$\xi_l^k = \frac{P_l^k - P_l^{\rm true}}{P_l^{\rm true}}. \tag{27}$$

Fig. 4 demonstrates that HADES completes burn-in after ∼20 samples in the case of the fiducial calculation (right-hand panels). However, the burn-in history for the full observation (left-hand panels) reveals a more interesting behaviour.

Initially, the power spectra show huge excursions at large scales. This is due to the observational mask and the fact that initially these regions are dominated by the constant initial guess ($r_k^0 = 1$). It is interesting to note that the first sample seems to be much closer to the true underlying power spectrum at the smaller scales, while the 20th sample is much further away. This clearly demonstrates the non-linear behaviour of the lognormal Poissonian sampler. We observe that with iterative correction of the large-scale power, the entire power spectrum progressively approaches the true mock power spectrum. This can be seen nicely in the lower left-hand panel of Fig. 4. After 100 samples have been calculated the true mock power spectrum is recovered for all the following samples. Thus, the initial burn-in period for a realistic observational setting can be expected to be of the order of 100 samples. Such a burn-in period is numerically not very demanding, and can easily be achieved in even higher resolution calculations.

Further, we ran a full Markov analysis for both test cases, by calculating 20 000 samples with a resolution of 64^3 voxels. We then estimate the ensemble mean and compared the recovered density field in the observed region via a point-to-point statistic to the true underlying mock signal. The results are presented in the upper panels of Fig. 5. It can be seen that both results are strongly correlated with the true underlying signal. To emphasize this fact, we also calculate the correlation factor given as

$$c = \frac{\sum_{i=0}^{N-1} \delta_i^{\rm true}\, \delta_i^{\rm mean}}{\sqrt{\sum_{i=0}^{N-1} \left(\delta_i^{\rm true}\right)^2}\sqrt{\sum_{i=0}^{N-1} \left(\delta_i^{\rm mean}\right)^2}}. \tag{28}$$

The correlation factors for the two test scenarios are also given in Fig. 5. They clearly demonstrate that the Hamiltonian sampler was able to recover the underlying density field to high accuracy in both cases.

7.3 Convergence

Testing the convergence of Markov chains is subject of many discussions in literature (see e.g. Heidelberger & Welch 1981; Gelman & Rubin 1992; Geweke 1992; Raftery & Lewis 1995; Cowles & Carlin 1996; Hanson 2001; Dunkley et al. 2005). In principle, there exist two categories of possible diagnostics. The methods of the first category rely on comparing interchain quantities between several chains while others try to estimate the convergence behaviour from interchain quantities within a single chain. In this paper we use the widely used Gelman & Rubin diagnostic, which is based on multiple simulated chains by comparing the variances within each

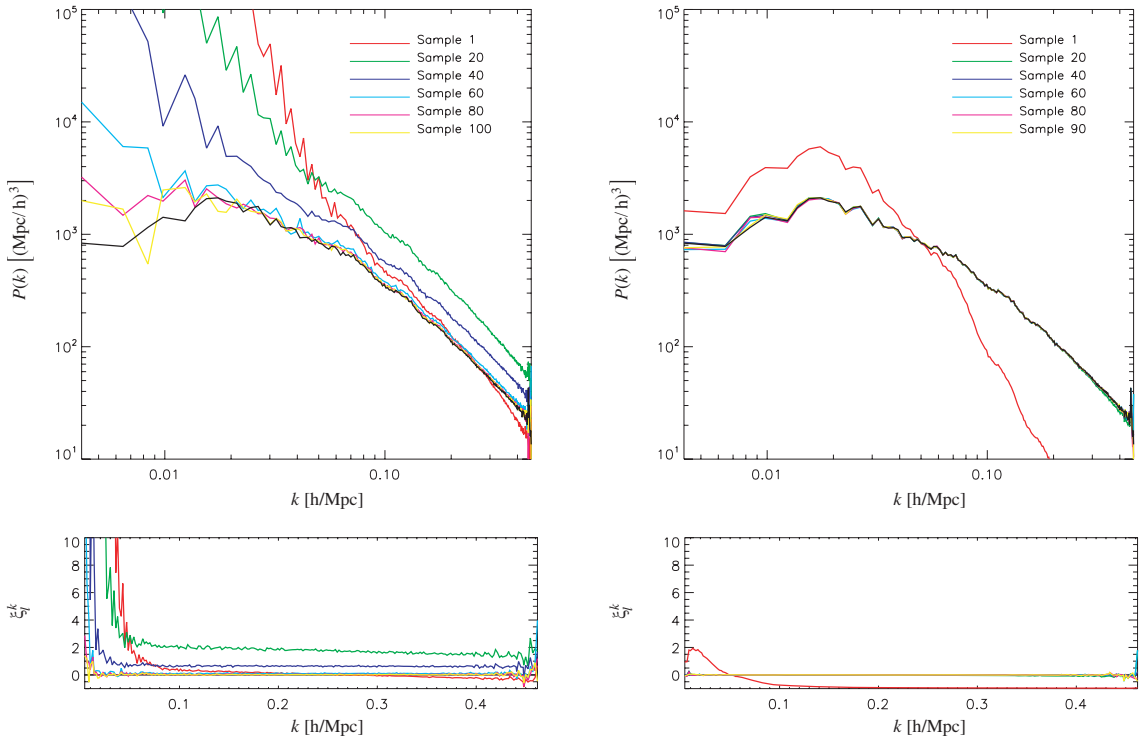

Figure 4. Successive power spectra measured from the Hamiltonian samples during burn-in. The right-hand panels correspond to the fiducial calculation, while the left-hand panels display the burn-in behaviour of the complete observational problem. The upper panels show the convergence of the individual sample spectra towards the spectrum of the true underlying matter field realization (black curve). The lower panels display the deviation from the true underlying spectrum ξ_l^k, demonstrating good convergence at the end of the burn-in period.

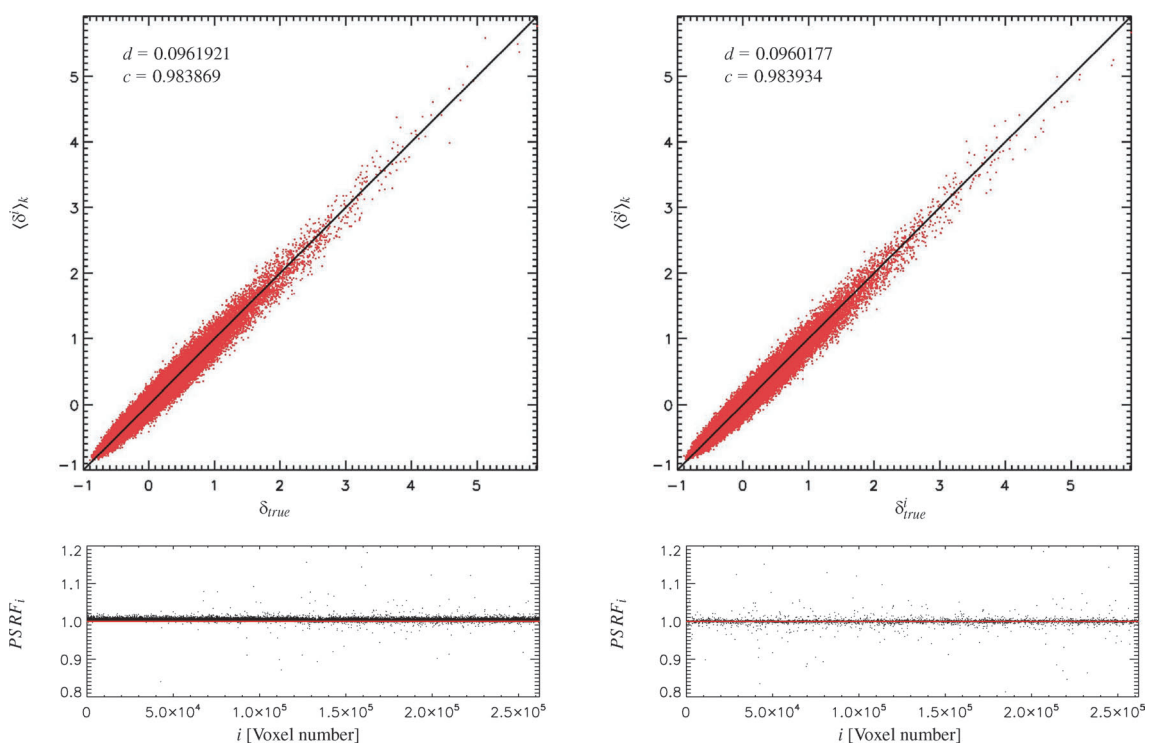

Figure 5. The upper panels show the point-to-point statistic of the ensemble mean field to the true underlying density field in the observed region for the fiducial calculation (right-hand panel) and the full observational problem (left-hand panel). The numbers in the upper left-hand part of the plots correspond to the Euclidean distance d and the correlation factor c. In the lower panels we plotted the results of the Gelman & Rubin convergence diagnostic for the corresponding tests. The PSRF indicates good convergence.

chain and the variance between chains (Gelman & Rubin 1992). In particular, we calculate the potential scale reduction factor (PSRF) (see Appendix B for details). A large PSRF indicates that the inter-chain variance is substantially greater than the intrachain variance and longer chains are required. Once the PSRF approaches unity, one can conclude that each chain has reached the target distribution.

We calculated the PSRF for each voxel of our test cases for chains with length $N_{\rm samp} = 20\,000$. The results for the two tests, as discussed above, are presented in Fig. 5. They clearly indicate the convergence of the Markov chains.

For the time being we use the Gelman & Rubin statistic to test convergence because of technical simplicity, although for the expense of having to calculate at least two chains. In the future we plan to explore other convergence diagnostics. In particular we are aiming at including intrachain methods as proposed in Hanson (2001) or Dunkley et al. (2005). This would allow us to detect convergence behaviour within the chain during burn-in. Such a convergence criterion could then be used to adjust the Hamiltonian masses for optimal sampling efficiency, as was proposed in Taylor et al. (2008).

7.4 Testing with simulated galaxy surveys

In this section, we describe the application of HADES to a mock galaxy survey based on the Millennium Run (Croton et al. 2006). The intention of this exercise is to test HADES in a realistic observational scenario. In particular, we want to demonstrate that HADES is able to reconstruct the fully evolved non-linear density field of the N-body simulation. The mock galaxy survey consists of a set of comoving galaxy positions distributed in a 500-Mpc box. To introduce survey geometry and selection effects, we virtually observe these galaxies through the sky mask and according to the selection function described in Section 7.1. The resulting galaxy distribution is then sampled to a 128^3 grid. This mock observation is then processed by HADES, which generates 20 000 lognormal Poissonian samples.

In Fig. 6 we present successive slices through density samples of the initial burn-in period. As can be seen, the first Hamiltonian sample (upper panels in Fig. 6) is largely corrupted by the false density information in the masked regions. This is due to the fact that the Hamiltonian sampler cannot be initialized with a point on the posterior surface. The initial samples are therefore required to identify a point on the corresponding posterior surface. As can be seen, the power in the unobserved and observed regions equalizes in the following samples. Also note that the first density sample depicts only very coarse structures. However, subsequent samples resolve finer and finer details. With the hundredth sample burn-in is completed. The lower panels of Fig. 6 demonstrate that the Hamiltonian sampler nicely recovers the filamentary structure of the density field.

Being a fully Bayesian method, the Hamiltonian sampler does not aim at calculating only a single estimate, such as a mean or maximum a posteriori value, it rather produces samples from the full lognormal Poissonian posterior. Given these samples we are able to calculate any desired statistical summary. In particular, we are able to calculate the mean and the variance of the Hamiltonian samples.

In Fig. 7 we show three different volume renderings of the ensemble mean density and the ensemble variance fields. It can be seen that the variance projections nicely reflect the Poissonian noise structure. Comparing high-density regions in the ensemble mean projections to the corresponding positions in the variance projections reveals a higher variance contribution for these regions, as expected for Poissonian noise. This demonstrates that our method allows us to provide uncertainty information for any resulting final estimate.

8 SUMMARY AND CONCLUSION

In this paper we introduced the HMC sampler for non-linear large-scale structure inference and demonstrated its performance in a variety of tests. As already described above, according to observational evidence and theoretical considerations, the posterior for non-linear density field inference is adequately represented by a lognormal Poissonian distribution, up to overdensities of $\delta \sim 100$. Hence, any method aiming at precision estimation of the fully evolved large-scale structure in the Universe needs to handle the non-linear relation between observations and the signal we seek to recover. The HMC sampler, presented in this paper, is a fully Bayesian method, and as such tries to evaluate the lognormal Poissonian posterior, given in equation (5), via sampling. In this fashion, the scientific output of the method is not a single estimate, but a sampled representation of the multidimensional posterior distribution. Given this representation of the posterior any desired statistical summary, such as mean, mode or variances can easily be calculated. Further, any uncertainty can seamlessly be propagated to the finally estimated quantities, by simply applying the corresponding estimation procedure to all Hamiltonian samples.

Unlike conventional Metropolis–Hastings algorithms, which move through the parameter space by random walk, the HMC sampler suppresses random walk behaviour by following a persistent motion. The HMC exploits techniques developed to follow classical dynamical particle motion in potentials, which, in the absence of numerical errors, yield an acceptance probability of unity. Although, in this paper we focused on the use of the lognormal Poissonian posterior, the method is more general. The discussion of the Hamiltonian sampler in Section 4 demonstrates that the method can in principle take into account a broad class of posterior distributions.

In Section 7, we demonstrated applications of the method to mock test cases, taking into account observational uncertainties such as selection effects, survey geometries and noise. These tests were designed to study the performance of the method in real-world applications.

In particular, it was of interest to establish intuition for the behaviour of the Hamiltonian sampler during the initial burn-in phase. Especially, the required amount of samples before the sampler starts drawing samples from the correct posterior distribution was of practical relevance. The tests demonstrated that for a realistic set-up, the initial burn-in period is of the order of $\sim$100 samples.

Further, the tests demonstrated that the Hamiltonian sampler produces unbiased samples, in the sense that each sample possesses correct power. Unlike a filter, which suppresses the signal in low signal-to-noise ratio regions, the Hamiltonian sampler non-linearly augments the poorly or unobserved regions with correct statistical information. In this fashion, each sample represents a complete matter field realization consistent with the observations.

The convergence of the Markov chain was tested via a Gelman & Rubin diagnostic. We compared the intrachain and interchain variances of two Markov chains each of length 20 000 samples. The estimated PSRF indicated good convergence of the chain. This result demonstrates that it is possible to efficiently sample from non-Gaussian distributions in very high-dimensional spaces.

In a final test the method was applied to a realistic galaxy mock observation based on the Millennium Run (Croton et al. 2006). Here we introduced again survey geometry and selection effects and

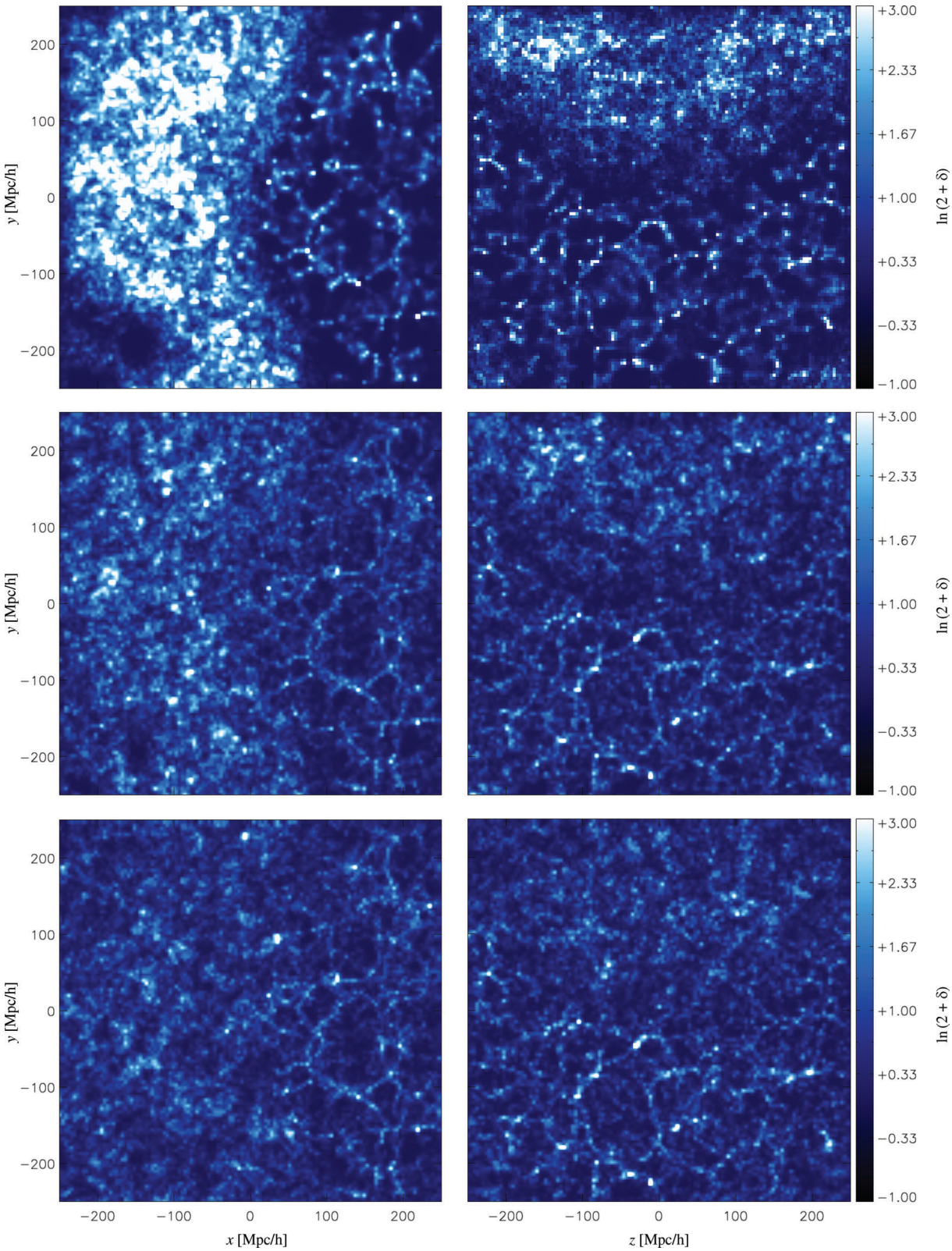

Figure 6. Slices through density samples during the initial burn-in phase. The upper panels correspond to the first sample, middle panels show the 10th sample and the lower panels present the 100th sample. Left- and right-hand panels show two different slices through the corresponding sample. It can be seen that during the initial burn-in phase power equalizes between the observed and unobserved regions. Successive samples recover finer and finer details.

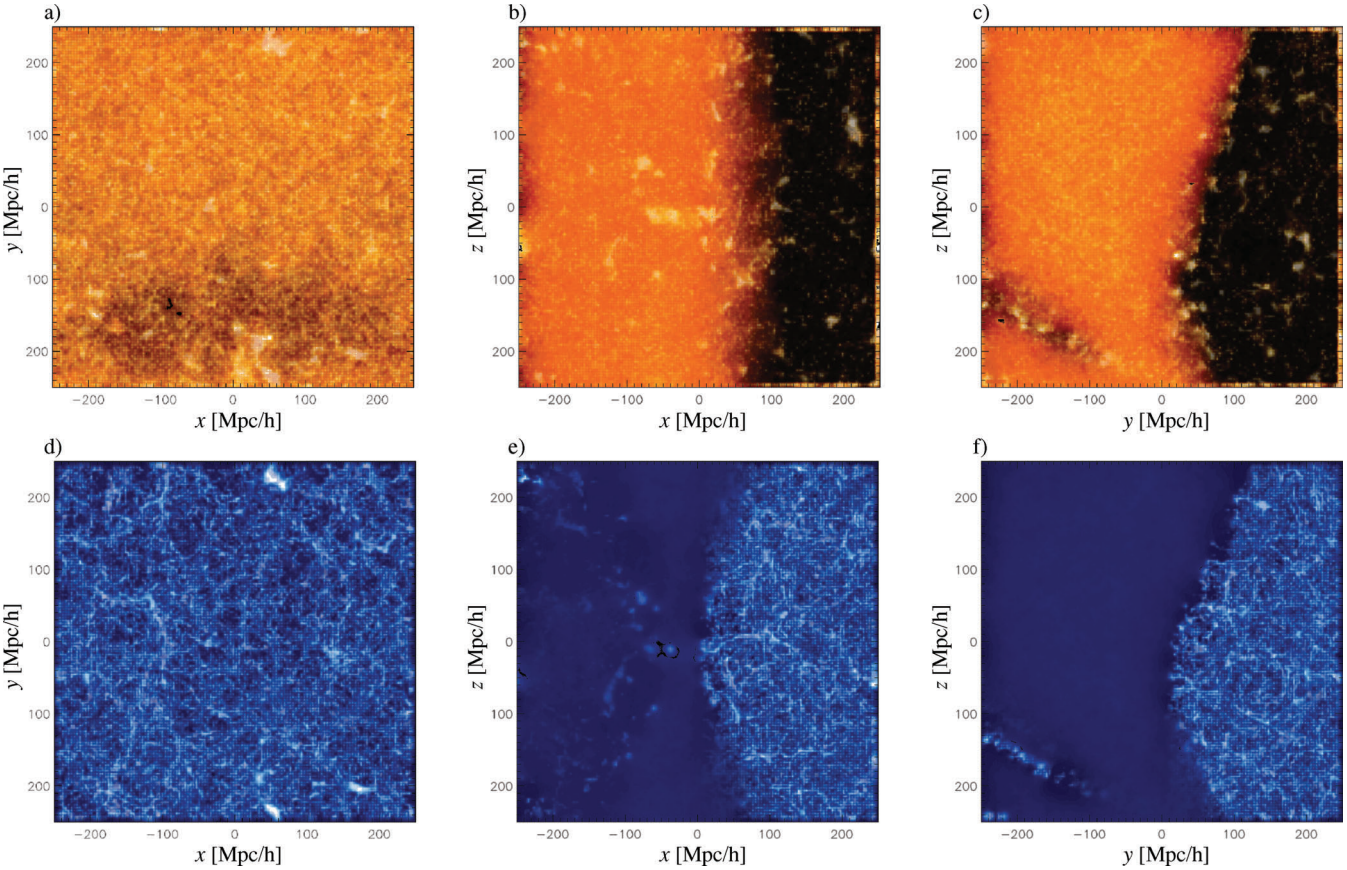

Figure 7. Volume rendering of the ensemble variance (upper panels) and the ensemble mean (lower panels) obtained from the mock galaxy catalogue analysis for three different perspectives. The mean shows filigree structures which have been recovered. It can also be seen that the ensemble variance reflects the Poissonian behaviour of the noise. High-density regions in the ensemble mean field correspond to regions with high variance as is expected for a Poissonian shot-noise contribution.

generated 20 000 samples of the lognormal Poissonian posterior. The results nicely demonstrate that the Hamiltonian sampler recovers the filamentary structure of the underlying matter field realization. For this test we also calculated the ensemble mean and the corresponding ensemble variance of the Hamiltonian samples, demonstrating that the Hamiltonian sampler also provides error information for a final estimate.

To conclude, in this paper we present a new and numerically efficient Bayesian method for large-scale structure inference and its numerical implementation HADES. HADES provides a sampled representation of the very high-dimensional non-Gaussian large-scale structure posterior, conditional on galaxy observations. This permits us to easily calculate any desired statistical summary, such as mean, mode and variance. In this fashion HADES is able to provide uncertainty information to any final quantity estimated from the Hamiltonian samples. The method, as presented here, is very flexible and can easily be extended to take into account additional non-linear observational constraints and joint uncertainties.

In summary, HADES, in its present form, provides the basis for future non-linear high-precision large-scale structure analysis.

ACKNOWLEDGMENTS

We would like to thank Rainer Moll and Björn Malte Schäfer for useful discussions and support with many valuable numerical gadgets. Further, we thank R. Benton Metcalf for useful remarks and commentaries, and Torsten A. Enß lin for encouraging us to pursue this project. Special thanks also to María Ángeles Bazarra-Castro for helpful assistance during the course of this project. We also thank the 'Transregional Collaborative Research Centre TRR 33 – The Dark Universe' for the support of this work.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Andrieu C., de Freitas N., Doucet A., Jordan M. I., 2003, Machine Learning, 50, 5
Bertschinger E., Dekel A., 1989, ApJ, 336, L5
Bertschinger E., Dekel A., 1991, in Latham D. W., da Costa L. A. N., eds, ASP Conf. Ser. Vol. 15, Large-Scale Structures and Peculiar Motions in the Universe. Astron. Soc. Pac., San Francisco, p. 67
Bistolas V., Hoffman Y., 1998, ApJ, 492, 439
Coles P., Jones B., 1991, MNRAS, 248, 1
Cowles M. K., Carlin B. P., 1996, J. American Statistical Association, 91, 883
Croton D. J. et al., 2006, MNRAS, 365, 11
Duane S., Kennedy A. D., Pendleton B. J., Roweth D., 1987, Phys. Lett. B, 195, 216
Dunkley J., Bucher M., Ferreira P. G., Moodley K., Skordis C., 2005, MNRAS, 356, 925
Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
Eisenstein D. J., Hu W., 1999, ApJ, 511, 5
Enßlin T. A., Frommert M., Kitaura F. S., 2009, Phys. Rev. D, 80, 105005
Erdoğdu P. et al., 2004, MNRAS, 352, 939
Fisher K. B., Lahav O., Hoffman Y., Lynden-Bell D., Zaroubi S., 1995, MNRAS, 272, 885

Frigo M., Johnson S. G., 2005, Proc. IEEE, 93, 216
Galassi M., Davies J., Theiler J., Gough B., Jungman G., Booth M., Rossi F., 2003, GNU Scientific Library: Reference Manual. Network Theory Ltd.
Gaztanaga E., Yokoyama J., 1993, ApJ, 403, 450
Gelman A., Rubin D., 1992, Statistical Sci., 7, 457
Geweke J., 1992, Bayesian Statistics. University Press, p. 169
Hajian A., 2007, Phys. Rev. D, 75, 083525
Hanson K. M., 2001, in Sonka M., Hanson K. M., eds, Proc. SPIE Conf. Ser. Vol. 4322, Markov Chain Monte Carlo Posterior Sampling with the Hamiltonian Method. SPIE, Bellingham, WA, p. 456
Heidelberger P., Welch P. D., 1981, Communications ACM, 24, 233
Hoffman Y., 1994, in Balkowski C., Kraan-Korteweg R. C., eds, ASP Conf. Ser. Vol. 67, Unveiling large-scale structures behind the Milky Way. Astron. Soc. Pac., San Francisco, p. 185
Hubble E., 1934, ApJ, 79, 8
Jasche J., Kitaura F. S., Wandelt B. D., Enßlin T. A., 2010, MNRAS, in press (doi:10.1111/j.1365-2966.2010.16610.x) (arXiv:0911.2493)
Kayo I., Taruya A., Suto Y., 2001, ApJ, 561, 22
Kitaura F. S., Enßlin T. A., 2008, MNRAS, 389, 497
Kitaura F. S., Jasche J., Li C., Ensslin T. A., Metcalf R. B., Wandelt B. D., Lemson G., White S. D. M., 2009, MNRAS, 400, 183
Lahav O., Fisher K. B., Hoffman Y., Scharf C. A., Zaroubi S., 1994, ApJ, 423, L93
Layzer D., 1956, AJ, 61, 383
Martínez V. J., Saar E., 2002, Statistics of the Galaxy Distribution. Chapman & Hall, Boca Raton
Matsumoto M., Nishimura T., 1998, ACM Trans. Modeling Comput. Simulation, 8, 3
Neal R. M., 1993, Technical Report CRG-TR-93-1, Probabilistic Inference Using Markov Chain Monte Carlo Methods. Department of Computer Science, Univ. Toronto
Neal R. M., 1996, Bayesian Learning for Neural Networks (Lecture Notes in Statistics), 1st edn. Springer-Verlag, New York
Peebles P. J. E., 1980, Large-scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Raftery A. E., Lewis S. M., 1995, in Gilks W. R., Spiegelhalter D. J., Richardson S., eds, Practical Markov Chain Monte Carlo. Chapman & Hall, London, p. 115
Robert C. P., Casella G., 2005, Monte Carlo Statistical Methods (Springer Texts in Statistics). Springer-Verlag New York, Inc., Secaucus, NJ
Saunders W., Ballinger W. E., 2000, in Kraan-Korteweg R. C., Henning P. A., Andernach H., eds, ASP Conf. Ser. Vol. 218, Mapping the hidden Universe: the Universe behind the Milky Way. Astron. Soc. Pac., San Francisco, p. 181
Schmoldt I. M. et al., 1999, ApJ, 118, 1146
Sheth R. K., 1995, MNRAS, 277, 933
Taylor J. F., Ashdown M. A. J., Hobson M. P., 2008, MNRAS, 389, 1284
Webster M., Lahav O., Fisher K., 1997, MNRAS, 287, 425
Zaroubi S., 2002, MNRAS, 331, 901

APPENDIX A: HAMILTONIAN MASSES

The Hamiltonian sampler can be extremely sensitive to the choice of masses. To estimate a good guess of Hamiltonian masses we follow a similar approach as suggested in Taylor et al. (2008). According to the leapfrog scheme, given in equations (20)–(22), a single application of the leapfrog method can be written in the form

$$p_i(t+\epsilon) = p_i(t) - \frac{\epsilon}{2}\left[\left.\frac{\partial\Psi(r)}{\partial r_i}\right|_{r(t)} + \left.\frac{\partial\Psi(r)}{\partial r_i}\right|_{r(t+\epsilon)}\right], \tag{A1}$$

$$r_i(t+\epsilon) = r_i(t) + \epsilon\sum_j \mathbf{M}^{-1}_{ij}\, p_j(t) - \frac{\epsilon^2}{2}\sum_j \mathbf{M}^{-1}_{ij}\left.\frac{\partial\Psi(r)}{\partial r_j}\right|_{r(t)}. \tag{A2}$$

We will then approximate the forces given in equation (17) for $r_i \ll 1$:

$$\begin{aligned}\frac{\partial\psi(r)}{\partial r_l} &= \sum_j \mathbf{Q}^{-1}_{lj}(r_j+\mu_j) - \left(\frac{N^{\rm g}_l}{1+b\,(\mathrm{e}^r-1)} - R_l\bar{N}\right) b\,\mathrm{e}^{r_l}\\ &\approx \sum_j \mathbf{Q}^{-1}_{lj}(r_j+\mu_j) - \left\{\left(N^{\rm g}_l - R_l\bar{N}\right) b\right.\\ &\quad \left. + \left[\left(N^{\rm g}_l - R_l\bar{N}\right) b - N^{\rm g}_l b^2\right] r_l\right\}\\ &= \sum_j \left\{\mathbf{Q}^{-1}_{lj} - \left[\left(N^{\rm g}_l - R_l\bar{N}\right) b - N^{\rm g}_l b^2\right]\delta^K_{lj}\right\} r_j\\ &\quad + \sum_j \mathbf{Q}^{-1}_{lj}\mu_j - \left(N^{\rm g}_l - R_l\bar{N}\right) b.\end{aligned} \tag{A3}$$

By introducing

$$\mathbf{A}_{lj} = \mathbf{Q}^{-1}_{lj} - \left[\left(N^{\rm g}_l - R_l\bar{N}\right) b - N^{\rm g}_l b^2\right]\delta^K_{lj} \tag{A4}$$

and

$$\mathbf{D}_l = \sum_j \mathbf{Q}^{-1}_{lj}\mu_j - \left(N^{\rm g}_l - R_l\bar{N}\right) b, \tag{A5}$$

equation (A3) simplifies to

$$\frac{\partial\psi(r)}{\partial r_l} = \sum_j \mathbf{A}_{lj}\, r_j + \mathbf{D}_l. \tag{A6}$$

Introducing this approximation into equations (A1) and (A2) yields

$$\begin{aligned}p_i(t+\epsilon) &= \sum_m \left[\delta^K_{im} - \frac{\epsilon^2}{2}\sum_j \mathbf{A}_{ij}\mathbf{M}^{-1}_{jm}\right] p_m(t)\\ &\quad - \epsilon\sum_j \mathbf{A}_{ij}\sum_p \left[\delta^K_{jp} - \frac{\epsilon^2}{4}\sum_m \mathbf{M}^{-1}_{jm}\mathbf{A}_{mp}\right] r_p(t)\\ &\quad - \frac{\epsilon}{2}\sum_m \left[\delta^K_{im} - \frac{\epsilon^2}{2}\sum_j \mathbf{A}_{ij}\mathbf{M}^{-1}_{jm}\right]\mathbf{D}_m\end{aligned} \tag{A7}$$

and

$$\begin{aligned}r_i(t+\epsilon) &= \epsilon\sum_j \mathbf{M}^{-1}_{ij}\, p_j(t)\\ &\quad + \sum_m \left(\delta^K_{im} - \frac{\epsilon^2}{2}\sum_j \mathbf{M}^{-1}_{ij}\,\mathbf{A}_{jm}\right) r_m(t)\\ &\quad - \frac{\epsilon^2}{2}\sum_j \mathbf{M}^{-1}_{ij}\mathbf{D}_j.\end{aligned} \tag{A8}$$

This result can be rewritten in matrix notation as

$$\begin{pmatrix} r(t+\epsilon)\\ p(t+\epsilon)\end{pmatrix} = \mathbf{T}\begin{pmatrix} r(t)\\ p(t)\end{pmatrix} - \frac{\epsilon^2}{2}\begin{pmatrix}\mathbf{M}^{-1}\,\mathbf{D}\\ \epsilon\left[\mathbf{I} - \frac{\epsilon^2}{2}\mathbf{A}\,\mathbf{M}^{-1}\right]\mathbf{D}\end{pmatrix}, \tag{A9}$$

where the matrix $\mathbf{T}$ is given as

$$\mathbf{T} = \begin{pmatrix}\left[\mathbf{I} - \frac{\epsilon^2}{2}\mathbf{M}^{-1}\mathbf{A}\right] & \epsilon\mathbf{M}^{-1}\\ -\epsilon\,\mathbf{A}\left[\mathbf{I} - \frac{\epsilon^2}{4}\mathbf{M}^{-1}\,\mathbf{A}\right] & \left[\mathbf{I} - \frac{\epsilon^2}{2}\mathbf{A}\,\mathbf{M}^{-1}\right]\end{pmatrix}, \tag{A10}$$

with $\mathbf{I}$ being the identity matrix. Successive applications of the leapfrog step yield the following propagation equation:

$$\begin{pmatrix} r^n\\ p^n\end{pmatrix} = \mathbf{T}^n\begin{pmatrix} r^0\\ p^0\end{pmatrix} - \frac{\epsilon^2}{2}\left[\sum_{i=0}^{n-1}\mathbf{T}^i\right]\begin{pmatrix}\mathbf{M}^{-1}\,\mathbf{D}\\ \epsilon\left[\mathbf{I} - \frac{\epsilon^2}{2}\mathbf{A}\,\mathbf{M}^{-1}\right]\mathbf{D}\end{pmatrix}. \tag{A11}$$

This equation demonstrates that there are two criteria to be fulfilled if the method is to be stable under repeated application of the leapfrog step. First, we have to ensure that the first term of equation (A11) does not diverge. This can be fulfilled if the eigenvalues of $\mathbf{T}$ have unit modulus. The eigenvalues λ are found by solving the characteristic equation

$$\det\left[\mathbf{I}\,\lambda^2 - 2\,\lambda\left(\mathbf{I} - \frac{\epsilon^2}{2}\mathbf{A}\,\mathbf{M}^{-1}\right) + \mathbf{I}\right] = 0. \tag{A12}$$

Note that this is a similar result to what was found in Taylor et al. (2008). Our aim is to explore the parameter space rapidly, and therefore we wish to choose the largest ϵ still compatible with the stability criterion. However, any dependence of equation (A12) also implies that no single value of ϵ will meet the requirement for every eigenvalue to have unit modulus. For this reason we choose

$$\mathbf{A} = \mathbf{M}. \tag{A13}$$

We then yield the characteristic equation

$$\left[\lambda^2 - 2\lambda\left(1 - \frac{\epsilon^2}{2}\right) + 1\right]^N = 0, \tag{A14}$$

where N is the number of voxels. This yields the eigenvalues

$$\lambda = \pm\mathrm{i}\sqrt{1 - \left[1 - \frac{\epsilon^2}{2}\right]^2} + \left[1 - \frac{\epsilon^2}{2}\right], \tag{A15}$$

which have unit modulus for $\epsilon \leq 2$. The second term in equation (A11) involves evaluation of the geometric series $\sum_{i=0}^{n-1}\mathbf{T}^i$. However, the geometric series for a matrix converges if and only if $|\lambda_i| < 1$ for each λ_i eigenvalue of $\mathbf{T}$. This clarifies that the non-linearities in the Hamiltonian equations generally do not allow for arbitrary large pseudo-time-steps ϵ. In addition, for practical purposes we usually restrict the mass matrix to the diagonal of equation (A4). For these two reasons, in practice, we choose the pseudo-time-step ϵ as large as possible while still obtaining a reasonable rejection rate.

APPENDIX B: GELMAN & RUBIN DIAGNOSTIC

The Gelman & Rubin diagnostic is a multichain convergence test (Gelman & Rubin 1992). It is based on analysing multiple Markov chains by comparing intrachain variances, within each chain, and interchain variances. A large deviation between these two variances indicates non-convergence of the Markov chain. Let $\{\phi^k\}$, where $k = 1, \ldots, N$, be the collection of a single Markov chain output. The parameter ϕ^k is the kth sample of the Markov chain. Here, for notational simplicity, we will assume ϕ to be single dimensional. To test convergence with the Gelman & Rubin statistic, one has to calculate M parallel MCMC chains, which are initialized from different parts of the target distribution. After discarding the initial burn-in samples, each chain is of length n. We can then label the outputs of various chains as ϕ_m^k, with $k = 1, \ldots, N$ and $m = 1, \ldots, M$. The interchain variance B can then be calculated as

$$B = \frac{n}{M-1}\sum_{m=1}^{M}(\theta_m - \Omega)^2, \tag{B1}$$

where θ_m is given as

$$\theta_m = \frac{1}{n}\sum_{k=1}^{n}\phi_m^k, \tag{B2}$$

and Ω as

$$\Omega = \frac{1}{M}\sum_{m=1}^{M}\theta_m. \tag{B3}$$

Then the intrachain variance can be calculated as

$$W = \frac{1}{M}\sum_{m=1}^{M}\Gamma_m^2, \tag{B4}$$

with

$$\Gamma_m^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(\phi_m^k - \theta_m\right)^2. \tag{B5}$$

With the above definition the marginal posterior variance can be estimated via

$$V = \frac{n-1}{n}W + \frac{M+1}{nM}B. \tag{B6}$$

If all M chains have reached the target distribution, this posterior variance estimate should be very close to the intrachain variance W. For this reason, one expects the ratio V/W to be close to 1. The square root of this ratio is referred to as the PSRF:

$$\mathrm{PSRF} = \sqrt{\frac{V}{W}}. \tag{B7}$$

If the PSRF is close to 1, one can conclude that each chain has stabilized, and has reached the target distribution (Gelman & Rubin 1992).

APPENDIX C: THE BLACKWELL–RAO ESTIMATOR AND THE BIAS

One particular advantage of the Hamiltonian sampling procedure is that it provides a sampled representation of the target distribution given in equation (5). The sampled distribution can then be expressed by a set of Dirac delta distributions (see e.g. Andrieu et al. 2003; Robert & Casella 2005):

$$\mathcal{P}_{N_{\mathrm{samp}}}\left(\{s_k\}|\left\{N_k^{\mathrm{g}}\right\}\right) = \frac{1}{N_{\mathrm{samp}}}\sum_{i=1}^{N_{\mathrm{samp}}}\delta^{\mathbf{D}}\left(\{s_k\} - \left\{s_k^i\right\}\right), \tag{C1}$$

where $\{s_k^i\}$ are the individual Hamiltonian samples. This allows for constructing various Blackwell–Rao estimators for many quantities in post-processing. In particular, we can provide the posterior distribution for biased density fields $\{\tilde{s}_k\}$ with a general non-linear, non-local biasing function $B(s)_k$ via the Blackwell–Rao estimator

$$\begin{aligned}\mathcal{P}_{N_{\mathrm{samp}}}\left(\{\tilde{s}_k\}|\left\{N_k^{\mathrm{g}}\right\}, B\right) &= \int \mathrm{d}\{s_k\}\,\mathcal{P}(\{\tilde{s}_k\}|\{s_k\}, B)\,\mathcal{P}\left(\{s_k\}|\left\{N_k^{\mathrm{g}}\right\}\right)\\ &\approx \int \mathrm{d}\{s_k\}\,\mathcal{P}(\{\tilde{s}_k\}|\{s_k\}, B)\,\mathcal{P}_{N_{\mathrm{samp}}}\left(\{s_k\}|\left\{N_k^{\mathrm{g}}\right\}\right)\\ &= \int \mathrm{d}\{s_k\}\,\mathcal{P}(\{\tilde{s}_k\}|\{s_k\}, B)\\ &\quad\times\frac{1}{N_{\mathrm{samp}}}\sum_{i=1}^{N_{\mathrm{samp}}}\delta^{\mathbf{D}}\left(\{s_k\} - \left\{s_k^i\right\}\right)\\ &= \frac{1}{N_{\mathrm{samp}}}\sum_{i=1}^{N_{\mathrm{samp}}}\mathcal{P}\left(\{\tilde{s}_k\}|\left\{s_k^i\right\}, B\right).\end{aligned} \tag{C2}$$

As can be seen the Blackwell–Rao estimate is obtained by simple marginalization over the density samples.

In the usual case that we assume a specific exact biasing relation, i.e.

$$\tilde{s}_k = B(s)_k, \tag{C3}$$

the result further simplifies considerably as the probability distribution $\mathcal{P}(\{\tilde{s}_k\}|\{s_k^i\}, B)$ collapses to a Dirac delta distribution. We can then write

$$\mathcal{P}_{N_{\rm samp}}\left(\{\tilde{s}_k\}|\left\{N_k^{\rm g}\right\}, B\right) = \frac{1}{N_{\rm samp}}\sum_{i=1}^{N_{\rm samp}} \mathcal{P}(\{\tilde{s}_k\}|\{s_k^i\}, B)$$
$$= \frac{1}{N_{\rm samp}}\sum_{i=1}^{N_{\rm samp}} \delta^{\mathbf{D}}\left(\{\tilde{s}_k\} - \left\{B(s^i)_k\right\}\right). \quad \text{(C4)}$$

In this fashion, many density posterior distributions with different bias models can be generated in a trivial post-processing step, by simply applying the biasing relation to all Hamiltonian samples. Note that this Blackwell–Rao estimator is a unique feature of the sampling approach since it relies on the set of Hamiltonian samples.

As one usually might want to investigate various different biasing models and since constructing the Blackwell–Rao estimates for the biased density fields requires only a negligible fraction of the total computation time required for the calculation of the entire Markov chain, the post-processing generation of Blackwell–Rao estimates is an ideal method to perform such studies. Also note that in this fashion all observational non-Gaussian uncertainties are propagated non-linearly to the finally inferred quantities.

This paper has been typeset from a TEX/LATEX file prepared by the author.