Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **418,** 729–746 (2011) doi:10.1111/j.1365-2966.2011.19525.x

What do cluster counts really tell us about the Universe?

Robert E. Smith[1,2★] and Laura Marian[2★]

[1]*Institute for Theoretical Physics, University of Zurich, Zurich CH 8037, Germany*
[2]*Argelander-Institute for Astronomy, Auf dem Hügel 71, D-53121 Bonn, Germany*

Accepted 2011 July 27. Received 2011 July 26; in original form 2011 June 9

ABSTRACT

We study the covariance matrix of the cluster mass function in cosmology. We adopt a two-line attack: first, we employ the counts-in-cells framework to derive an analytic expression for the covariance of the mass function. Secondly, we use a large ensemble of N-body simulations in the Λ cold dark matter framework to test this. Our theoretical results show that the covariance can be written as the sum of two terms: a Poisson term, which dominates in the limit of rare clusters; and a sample variance term, which dominates for more abundant clusters. Our expressions are analogous to those of Hu & Kravtsov for multiple cells and a single mass tracer. Calculating the covariance depends on: the mass function and bias of clusters, and the variance of mass fluctuations within the survey volume. The predictions show that there is a strong bin-to-bin covariance between measurements. In terms of the cross-correlation coefficient, we find $r \gtrsim 0.5$ for haloes with $M \lesssim 3 \times 10^{14}\,h^{-1}\,\mathrm{M}_{\odot}$ at $z = 0$. Comparison of these predictions with estimates from simulations shows excellent agreement. We use the Fisher matrix formalism to explore the cosmological information content of the counts. We compare the Poisson likelihood model, with the more realistic likelihood model of Lima & Hu, and all terms entering the Fisher matrices are evaluated using the simulations. We find that the Poisson approximation should only be used for the rarest objects, $M \gtrsim 5 \times 10^{14}\,h^{-1}\,\mathrm{M}_{\odot}$, otherwise the information content of a survey of size $V \sim 13.5\,h^{-3}\,\mathrm{Gpc}^3$ would be overestimated, resulting in errors that are nearly two times smaller. As an auxiliary result, we show that the bias of clusters, obtained from the cluster–mass cross-variance, is linear on scales $>50\,h^{-1}$ Mpc, whereas that obtained from the auto-variance is non-linear.

Key words: cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

The last decade of research in cosmology has largely been focused on devising probes to reveal the physical nature of dark energy and the origin of the accelerated expansion of the Universe. Among the most promising probes, as identified for example in Albrecht et al. (2006), are cluster counts.

From a theoretical perspective, the abundance of clusters per unit solid angle $\mathrm{d}\Omega$ is an integral of the mass function over mass M and volume element $\mathrm{d}V$:

$$\frac{\mathrm{d}N}{\mathrm{d}\Omega} = \int \mathrm{d}z \frac{\mathrm{d}V}{\mathrm{d}\Omega\,\mathrm{d}z} \int_{M_{\mathrm{th}}(z)} \mathrm{d}M n(M)\,, \quad (1)$$

where $M_{\mathrm{th}}(z)$ is a redshift-dependent mass detection threshold for the clusters and where the mass function is defined as the number of haloes per unit volume and unit mass, i.e. $n(M) = \mathrm{d}N/\mathrm{d}V/\mathrm{d}M$, with M the virial mass. For a wide range of cosmological models, $n(M)$ can be accurately predicted from the semi-analytical prescriptions based on the spherical or ellipsoidal collapse model (e.g. Press & Schechter 1974; Sheth & Tormen 1999, and see Section 3.4 for more details).

The mass function is primarily sensitive to the statistics of the initial conditions and to the amplitude σ_8 and shape of the matter power spectrum. The latter depends on the matter density of the Universe Ω_{m}, the Hubble parameter h, the spectral index of the primordial power spectrum n, and the dark energy equation of state $w \equiv P_w/\rho_w$, where P_w and ρ_w are the pressure and energy density of the dark energy. The volume element integral in the above equation renders the cluster counts even more sensitive to Ω_{m} and w. Measuring the cluster abundance at different redshifts can constrain a dynamical w and thus enable one to differentiate between a cosmological constant Λ and alternative dark energy scenarios such as quintessence (Wang & Steinhardt 1998), or dark energy inhomogeneities coupling to dark matter (Manera & Mota 2006).

For many decades, the study of clusters of galaxies has been a centrepiece for observational cosmology, which has produced many

★E-mail: res@physik.unizh.ch (RES); lmarian@astro.uni-bonn.de (LM)

important results and cosmological inferences. Currently, there are four observational strategies for detecting clusters: *X*-ray emission (see Borgani et al. 2001; Reiprich & Böhringer 2002; Allen et al. 2003; Schuecker et al. 2003; Henry 2004; Mantz et al. 2008, 2010; Vikhlinin et al. 2009, and references therein); optical emission (see Gladders et al. 2007; Rozo et al. 2010, and references therein); the Sunyaev–Zel'Dovich (SZ) effect (Sunyaev & Zeldovich 1972) , i.e. the up-scattering of CMB photons off hot electrons in the intracluster medium (see Vanderlinde & The SPT Collaboration 2010; Planck Collaboration et al. 2011; Muchovej et al. 2011; Sehgal & The ACT Collaboration 2011, and references therein) and weak gravitational lensing (see Schirmer et al. 2007; Abate et al. 2009; Israel et al. 2010, and references therein).

One of the most challenging aspects of deriving cosmological constraints from cluster counts is the fact that virial masses are not directly observable: a conversion is needed to translate observables such as flux, luminosity, temperature and SZ decrement into mass. The mass–observable relation is degenerate with cosmological parameters, as shown in Lima & Hu (2005), and can severely degrade the inferred constraints. Substantial progress has been made in calibrating the mass–observable relation in the recent years, through numerical simulations, or by comparing different methods against each other (Zhang et al. 2007, 2008; Okabe et al. 2010). Lima & Hu (2005) also proposed a self-calibration technique that uses the clustering of clusters to break the degeneracy between the uncertainties in the mass–observable relation and cosmological parameters.

Owing to observational challenges, the cluster studies mentioned earlier employ small numbers of massive clusters (at most a few hundreds, but in general a few tens) to constrain cosmology. In obtaining these constraints it is widely assumed that the likelihood function for the selected clusters follows the Poisson distribution. Whilst this assumption may be reasonable for the most massive clusters, $M \sim 10^{15}\,h^{-1}\,\mathrm{M}_{\odot}$, it will certainly fail at lower masses. Future surveys, such as eROSITA (Predehl et al 2010), LSST (2009), Euclid (Refregier et al. 2010), Pan-STARRS,[1] DES (2005), will be able to detect large samples of intermediate-mass clusters, $M \sim 10^{14}\,h^{-1}\,\mathrm{M}_{\odot}$. In order to make accurate inferences from these data, the cluster likelihood function will require a more complex statistical treatment, and in particular knowledge about the covariance matrix of the mass function.

This paper is driven by the following two questions: what is the covariance matrix for measurements of the mass function? How much are forecasted errors, which rely on the Poisson approximation, affected by more realistic modelling of the cluster likelihood function? The main theoretical tools that we shall employ to answer these questions will be the counts-in-cell formalism introduced by Peebles (1980) and further developed by Hu & Kravtsov (2003, hereafter HK03) and Lima & Hu (2004, hereafter LH04). We shall also compare the theoretical results obtained via this formalism to measurements obtained from a large ensemble of *N*-body simulations.

As this paper was nearing submission, a study by Valageas et al. (2011) was reported; this work explores related, but complementary, questions to those presented here.

This paper is structured in the following way. In Section 2, we review the counts-in-cells formalism and also the extension to the cluster likelihood developed by LH04; in Section 3 we derive the mass function covariance in a formal way; in Section 4 we describe the numerical simulations from which we measure the mass function covariance; in Section 5 we present a comparison between the measured and the predicted covariance; and in Section 6 we use the Fisher matrix formalism to estimate the impact that the full covariance matrix of the mass function has on cosmological constraints. Finally, in Section 7 we discuss and summarize our findings.

[1] http://pan-starrs.ifa.hawaii.edu

2 THEORETICAL BACKGROUND

2.1 The cellular model

In this section, we give a short description of the counts-in-cell formalism, used by HK03 to compute the linear theory sample variance of cluster counts, and by LH04 to estimate the impact of the latter on Fisher matrix predictions.

Consider some large cubical patch of the Universe of volume V_{μ} and containing N clusters that possess some distribution of masses. Let us subdivide this volume into a set of N_c equal cubical cells and the mass distribution into a set of N_m mass bins. Let the number of clusters in the ith cell and in the α^{th} mass bin be denoted as $N_{i,\alpha}$. We shall assume that the probability that the ith cell contains $N_{i,\alpha}$ clusters in the mass bin α is a Poisson process:

$$P(N_{i,\alpha}|m_{i,\alpha}) = \frac{m_{i,\alpha}^{N_{i,\alpha}}\exp(-m_{i,\alpha})}{N_{i,\alpha}!}. \tag{2}$$

For any quantity X, we denote the average over the sampling distribution – the Poisson process in this case – as $\langle X\rangle_P$, and the ensemble average over many realizations of the density field as $\overline{X} \equiv \langle X\rangle_s$, termed sample variance in HK03. The average of $N_{i,\alpha}$ over the sampling distribution can be written as (see also Cole & Kaiser 1989; Mo & White 1996):

$$m_{i,\alpha} \equiv \overline{m}_{i,\alpha}\left[1 + \overline{b}_{\alpha}\delta_V(\boldsymbol{x}_i)\right], \tag{3}$$

where $\overline{m}_{i,\alpha} = \overline{n}_{\alpha}V_i$ is the ensemble- and Poisson-averaged number of counts in cell i and mass bin α. The volume of the cell and the cell-averaged overdensity are given by:

$$V_i = \int \mathrm{d}^3\boldsymbol{x}\; W(\boldsymbol{x}|\boldsymbol{x}_i); \tag{4}$$

$$\delta_V(\boldsymbol{x}_i) = \frac{1}{V_i}\int \mathrm{d}^3\boldsymbol{x}\; W(\boldsymbol{x}|\boldsymbol{x}_i)\delta(\boldsymbol{x}), \tag{5}$$

where $W(\boldsymbol{x}|\boldsymbol{x}_i)$ is the window function for the ith cell (see Section 3.2 for more details). The number density and linear bias of the clusters averaged over the mass bin α are given by:

$$\overline{n}_{\alpha} = \int_{M_{\alpha}-\Delta M_{\alpha}/2}^{M_{\alpha}+\Delta M_{\alpha}/2} \mathrm{d}M n(M)\,; \tag{6}$$

$$\overline{b}_{\alpha} = \frac{1}{\overline{n}_{\alpha}}\int_{M_{\alpha}-\Delta M_{\alpha}/2}^{M_{\alpha}+\Delta M_{\alpha}/2} \mathrm{d}M b(M)n(M), \tag{7}$$

where $b(M)$ is the linear bias of haloes of mass M.

As was shown in HK03, the correlations in the underlying density field induce a correlation in the number counts of the cells, defined as:

$$\begin{aligned} S_{ij}^{\alpha\beta} &\equiv \left\langle \left(N_{i,\alpha} - \overline{m}_{i,\alpha}\right)\left(N_{j,\beta} - \overline{m}_{j,\beta}\right)\right\rangle_{P,s} \\ &= \left\langle \left(m_{i,\alpha} - \overline{m}_{i,\alpha}\right)\left(m_{j,\beta} - \overline{m}_{j,\beta}\right)\right\rangle_{s} \\ &= \overline{m}_{i,\alpha}\overline{m}_{j,\beta}\overline{b}_{\alpha}\overline{b}_{\beta}\int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}W_i^*(\boldsymbol{k})W_j(\boldsymbol{k})P(\boldsymbol{k}), \end{aligned} \tag{8}$$

where for independent Poisson processes the probability $P(N_{i,\alpha}, N_{j,\beta}|m_{i,\alpha}, m_{j,\beta}) = P(N_{i,\alpha}|m_{i,\alpha})P(N_{j,\beta}|m_{j,\beta})$. In the last line we introduced the power spectrum $P(k)$ as the Fourier transform of the correlation function ξ,

$$\xi(\boldsymbol{r}) \equiv \langle \delta(\boldsymbol{x}_i)\delta(\boldsymbol{x}_j)\rangle_s = \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} P(\boldsymbol{k}) \exp\left(-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}\right). \quad (9)$$

$W_i(\boldsymbol{k})$ is the Fourier transform of the cell window function and $\boldsymbol{r} = \boldsymbol{x}_i - \boldsymbol{x}_j$ (see Section 3.2).

2.2 The Gauss–Poisson likelihood function for counts in cells

The likelihood of drawing a particular set of cluster counts $\boldsymbol{N} \in \{N_{1,1}, \ldots, N_{N_c,1}, N_{1,2}, \ldots, N_{N_c,N_m}\}$ in the cells, given a model for the counts in the cells $\overline{\boldsymbol{m}} \in \{\overline{m}_{1,1}, \ldots, \overline{m}_{N_c,1}, \overline{m}_{N_c,2} \ldots, \overline{m}_{N_c,N_m}\}$, was written by LH04:

$$\mathcal{L}(\boldsymbol{N}|\overline{\boldsymbol{m}}, \mathbf{S}) = \int \mathrm{d}^{\mathcal{N}}m \left[\prod_{\alpha=1}^{N_m}\prod_{i=1}^{N_c} P(N_{i,\alpha}|m_{i,\alpha})\right] G(\boldsymbol{m}|\overline{\boldsymbol{m}}, \mathbf{S}), \quad (10)$$

with $\mathcal{N} = N_c \times N_m$, and where it was assumed that the statistics of the cell-averaged density field are described by a multivariate Gaussian:

$$G(\boldsymbol{m}|\overline{\boldsymbol{m}}, \mathbf{S}) \equiv \frac{(2\pi)^{-N/2}}{|S|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{m}-\overline{\boldsymbol{m}})^{\mathrm{T}}\mathbf{S}^{-1}(\boldsymbol{m}-\overline{\boldsymbol{m}})\right], \quad (11)$$

with $\mathbf{S}$ defined in equation (8). Measurements of the bispectrum of the CMB have shown that the statistics of the initial fluctuations are very nearly Gaussian (Komatsu et al. 2010). Whilst we know that the non-linear growth of structure in the present epoch drives the statistics of the density field to become non-Gaussian, in the limit that the cells are large compared to the coherence length of the field, we expect that the Gaussian approximation will be very good.

At this point we may also be more precise about what we mean by ensemble and Poisson averages:

$$\langle X(\boldsymbol{N})\rangle_{P,s} \equiv \sum_{N_{1,1}=0}^{\infty} \ldots \sum_{N_{N_c,N_m}=0}^{\infty} \mathcal{L}(\boldsymbol{N}|\overline{\boldsymbol{m}}, \mathbf{S}) X(\boldsymbol{N}). \quad (12)$$

Equation (10) can be simplified in two limits.

(i) *Case I*: in the limit that the ensemble average variance is much smaller than the Poisson variance: i.e. $S_{ii} \ll \overline{m}_i$. In this case, the Gaussian effectively becomes a delta function centred on $\overline{\boldsymbol{m}}$ and the likelihood simply becomes a product of Poisson probabilities:

$$\mathcal{L}(\boldsymbol{N}|\overline{\boldsymbol{m}}) \approx \prod_{\alpha=1}^{N_m}\prod_{i=1}^{N_c} P(N_{i,\alpha}|\overline{m}_{i,\alpha}). \quad (13)$$

(ii) *Case II*: in the limit that the number of counts in each cell and mass bin is large, then the Poisson process becomes a Gaussian:

$$\prod_{\alpha=1}^{N_m}\prod_{i=1}^{N_c} P(N_{i,\alpha}|m_{i,\alpha}) \approx G(\boldsymbol{N}|\boldsymbol{m}, \mathbf{M}), \quad (14)$$

where $\mathbf{M} \rightarrow M^{ij}_{\alpha\beta} = \delta^K_{i,j}\delta^K_{\alpha,\beta}m_{i,\alpha}$. Hence, as shown in LH04, the likelihood function becomes:

$$\mathcal{L}(\boldsymbol{N}|\overline{m}, \mathbf{S}) \approx \int d^{\mathcal{N}}m G(\boldsymbol{N}|\boldsymbol{m}, \mathbf{M}) G(\boldsymbol{m}|\overline{\boldsymbol{m}}, \mathbf{S}) \quad (15)$$

and via the convolution theorem this can be approximated as a Gaussian with shifted mean and augmented covariance matrix:

$$\mathcal{L}(\boldsymbol{N}|\overline{\boldsymbol{m}}, \mathbf{S}) \approx G(\boldsymbol{N}|\overline{\boldsymbol{m}}, \mathbf{C}); \quad \mathbf{C} = \overline{\mathbf{M}} + \mathbf{S}, \quad (16)$$

where $\overline{\mathbf{M}} \rightarrow \overline{M}^{ij}_{\alpha\beta} = \delta^K_{i,j}\delta^K_{\alpha,\beta}\overline{m}_{i,\alpha}$. Note that in the above equation, the approximate sign is used since negative number counts are formally forbidden (for a more detailed discussion of this see Hu & Cohn 2006).

3 COVARIANCE OF THE MASS FUNCTION

The final result of Section 2 is that in the limit of a large number of counts per cell, the joint likelihood for all the cells is a Gaussian with model mean $\overline{\boldsymbol{m}}$ and with a covariance matrix, $\mathbf{C} = \overline{\mathbf{M}} + \mathbf{S}$. In the following section, we shall use these results to answer the question: what is the covariance matrix for measurements of the mass function?

3.1 A formal approach

The mass function $n(M)$ is the number density of clusters in a volume V, per unit mass. Using our counts in cells distribution, an estimator for the mass function in the i^{th} cell is:

$$\hat{n}_i(M_\alpha) = \frac{N_{i,\alpha}}{V_i \Delta M_\alpha}, \quad (17)$$

which, if we average over all cells and all cells have equal volume, becomes:

$$\hat{n}(M_\alpha) = \frac{1}{V_\mu \Delta M_\alpha}\sum_i N_{i,\alpha}. \quad (18)$$

The above estimate is unbiased, and its expectation value $\overline{n}(M_\alpha) \equiv \langle \hat{n}(M_\alpha)\rangle_{P,s}$ can be formally calculated using equation (12):

$$\begin{aligned}
\overline{n}(M_\alpha) &= \sum_{N_{1,1}=0}^{\infty} \ldots \sum_{N_{N_c,N_m}=0}^{\infty} \mathcal{L}(\boldsymbol{N}|\overline{\boldsymbol{m}}, \mathbf{S}) \sum_i \frac{N_{i,\alpha}}{V_\mu \Delta M_\alpha} \\
&= \int d^{\mathcal{N}}m G(\boldsymbol{m}|\overline{\boldsymbol{m}}, \mathbf{S}) \sum_{N_{1,1}=0}^{\infty} P(N_{1,1}|m_{1,1}) \ldots \\
&\quad \times \sum_{N_{N_c,N_m}=0}^{\infty} P\left(N_{N_c,N_m}|m_{N_c,N_m}\right) \sum_{i=1}^{N_c} \frac{N_{i,\alpha}}{V_\mu \Delta M_\alpha} \\
&= \frac{1}{V_\mu \Delta M_\alpha} \int d^{\mathcal{N}}m G(\boldsymbol{m}|\overline{\boldsymbol{m}}, \mathbf{S}) \sum_i m_{i,\alpha} \\
&= \sum_{i=1}^{N_c} \frac{\overline{m}_{i,\alpha}}{V_\mu \Delta M_\alpha}.
\end{aligned} \quad (19)$$

In a similar fashion, the covariance matrix of the cluster mass function can also be calculated:

$$\begin{aligned}
\mathcal{M}_{\alpha\beta} &\equiv \langle [n(M_\alpha) - \overline{n}(M_\alpha)]\,[n(M_\beta) - \overline{n}(M_\beta)]\rangle_{s,P} \\
&= \sum_{i,j} \frac{\langle N_{i,\alpha}N_{j,\beta}\rangle_{s,P}}{V_\mu^2 \Delta M_\alpha \Delta M_\beta} - \overline{n}(M_\alpha)\overline{n}(M_\beta),
\end{aligned} \quad (20)$$

where the expectation of the product of the counts can be written as:

$$\begin{aligned}
\sum_{i,j}\langle N_{i,\alpha}N_{j,\beta}\rangle_{s,P} &= \sum_{N_{1,1}=0}^{\infty} \ldots \sum_{N_{N_c,N_m}=0}^{\infty} \mathcal{L}(\boldsymbol{N}|\overline{\boldsymbol{m}}, \mathbf{S}) \sum_{i,j} N_{i,\alpha}N_{j,\beta} \\
&= \int d^{\mathcal{N}}m G(\boldsymbol{m}|\overline{\boldsymbol{m}}, \mathbf{S}) \left[\sum_{i,j,i\neq j\cup\alpha\neq\beta} m_{i,\alpha}m_{j,\beta} + \sum_i \left\langle N^2_{i,\alpha}\right\rangle_P\right].
\end{aligned} \quad (21)$$

Recall that $\overline{m}_{i,\alpha} = \overline{n}(M_\alpha)\Delta M_\alpha V_i$ and that for the Poisson distribution we have: $\langle X^2\rangle = \langle X\rangle[1 + \langle X\rangle]$. On inserting these relations into

the above equation, and on completing the sums, we find:

$$\sum_{i,j}\langle N_{i,\alpha}N_{j,\beta}\rangle_{s,P} = \int \mathrm{d}^{\mathcal{N}}m G(\boldsymbol{m}|\overline{\boldsymbol{m}},\mathbf{S}) \times \sum_{i,j}\left[m_{i,\alpha}m_{j,\beta}+m_{i,\alpha}\delta^K_{i,j}\delta^K_{\alpha,\beta}\right] = \sum_{ij}\left[S^{\alpha\beta}_{ij}+\overline{m}_{i,\alpha}\overline{m}_{j,\beta}+\overline{m}_{i,\alpha}\delta^K_{i,j}\delta^K_{\alpha,\beta}\right], \quad (22)$$

where in the last line we used equation (8). On inserting this result back into equation (20), we obtain:

$$\mathcal{M}_{\alpha\beta} = \sum_{ij}\frac{\left[\overline{m}_{i,\alpha}\delta^K_{i,j}\delta^K_{\alpha,\beta}+S^{\alpha\beta}_{ij}\right]}{{V_\mu}^2\Delta M_\alpha\Delta M_\beta} = \frac{\delta^K_{\alpha,\beta}\overline{n}(M_\alpha)}{V_\mu\Delta M_\alpha}+\frac{\overline{n}(M_\alpha)\overline{n}(M_\beta)\overline{b}_\alpha\overline{b}_\beta}{{V_\mu}^2} \times\sum_{ij}V_iV_j\int\frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}W_i^*(\boldsymbol{k})W_j(\boldsymbol{k})P(k). \quad (23)$$

Considering the first term in the above, we may simplify this expression by performing the sums over i and j, and the window functions i.e.

$$\sum_i V_iW_i(\boldsymbol{k}) = \sum_i V_i\int\mathrm{d}^3\boldsymbol{x}\exp\left[\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}\right]W(\boldsymbol{x}|\boldsymbol{x}_i) = \int\mathrm{d}^3\boldsymbol{x}\exp\left[\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}\right]\sum_i V_iW(\boldsymbol{x}|\boldsymbol{x}_i) = V_\mu\widetilde{W}(\boldsymbol{k}). \quad (24)$$

Hence, the covariance matrix can be written as:

$$\mathcal{M}_{\alpha\beta} = \overline{n}(M_\alpha)\overline{n}(M_\beta)\overline{b}_\alpha\overline{b}_\beta\sigma^2(V_\mu)+\frac{\delta^K_{\alpha,\beta}\overline{n}(M_\alpha)}{V_\mu\Delta M_\alpha}, \quad (25)$$

where $\sigma^2(V_\mu)$ is the mass density variance in the entire volume:

$$\sigma^2(V_\mu)\equiv\int\frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}|\widetilde{W}(\boldsymbol{k})|^2P(k). \quad (26)$$

From equation (25) it can be seen that the crucial quantity which controls the covariance between estimates of the mass function in different mass bins is $\sigma(V_\mu)$. The strength of the covariance is also modulated by the linear bias and the mass function in each of the bins considered.

3.2 A short-cut to the covariance

Whilst in the above we have presented a formal derivation of the mass function covariance from the HK03 and LH04 formalism, there is a more intuitive approach to arriving at the same result as given by equation (25), which we mention below.

Let us consider the limiting case where we have a single cell that fills the whole of our sample space $V_i \to V_\mu$; also $m_{i,\alpha}\to m_\alpha$ and similar for all the other quantities defined in the cells. The above formalism still applies, and we have that the covariance matrix of mass function can be written as:

$$\mathcal{M}_{\alpha\beta} = \frac{S^{\alpha,\beta}_{ii}}{V_\mu^2\Delta M_\alpha\Delta M_\beta}+\delta^K_{\alpha,\beta}\frac{m_{i,\alpha}}{V_\mu^2\Delta M_\alpha\Delta M_\beta} = \overline{n}(M_\alpha)\overline{n}(M_\beta)\overline{b}_\alpha\overline{b}_\beta\sigma^2(V_\mu)+\delta^K_{\alpha,\beta}\frac{n(M_\alpha)}{V_\mu\Delta M_\alpha}, \quad (27)$$

where $\sigma(V_\mu)$ is the variance in the total volume.

3.3 The cross-correlation coefficient

As a direct corollary to the previous results, we may write an expression for the correlation matrix, which is defined as:

$$r_{\alpha\beta}\equiv\frac{\mathcal{M}_{\alpha\beta}}{\sqrt{\mathcal{M}_{\alpha\alpha}\mathcal{M}_{\beta\beta}}}. \quad (28)$$

On factoring out $[\overline{n}(M_\alpha)/V_\mu\Delta M_\alpha]^{1/2}$ from $\sqrt{\mathcal{M}_{\alpha\alpha}}$ in the denominator, and a similar term from $\sqrt{\mathcal{M}_{\beta\beta}}$, and on using the fact that $\overline{m}_\alpha=\overline{n}(M_\alpha)\Delta M_\alpha V_\mu$, we find:

$$r_{\alpha\beta} = \frac{\sqrt{\overline{m}_\alpha\overline{m}_\beta}\,\overline{b}_\alpha\overline{b}_\beta\sigma^2(V_\mu)+\delta^K_{\alpha\beta}}{\left[1+\overline{m}_\alpha\overline{b}_\alpha^2\sigma^2(V_\mu)\right]^{1/2}\left[1+\overline{m}_\beta\overline{b}_\beta^2\sigma^2(V_\mu)\right]^{1/2}}. \quad (29)$$

Two limits are apparent: when $\sqrt{\overline{m}_\alpha\overline{m}_\beta}\,\overline{b}_\alpha\overline{b}_\beta\sigma^2(V_\mu)\ll 1$, then $r_{\alpha\beta}\to\delta^K_{\alpha,\beta}$ and the mass function covariance matrix is decorrelated; this would happen for the case of rare haloes, for which the mass function is very small. On the other hand, when $\sqrt{\overline{m}_\alpha\overline{m}_\beta}\,\overline{b}_\alpha\overline{b}_\beta\sigma^2(V_\mu)\gg 1$, then $r_{\alpha\beta}\to 1$ and the covariance matrix is fully correlated. This would be the case for smaller haloes, for which the mass function is quite large.

Finally, we note that taking $V_\mu\to\infty$ and hence $\sigma(V_\mu)\to 0$ *does not* guarantee that the correlation between different mass bins is negligible. As the above clearly shows, it is the quantity $V_\mu\sigma^2(V_\mu)$ that is required to vanish for negligible correlation to occur. For a power-law power spectrum, we would have that $V_\mu\sigma^2(V_\mu)\propto R^3R^{-(3+n)}\propto R^{-n}$, which only vanishes for $n>0$. For cold dark matter we have a rolling spectral index, and $n>0$ for $k\lesssim 0.01\,h\,\mathrm{Mpc}^{-1}$, which implies that $L_{\rm box}\gtrsim 500\,h^{-1}$ Mpc for the covariance to diminish.

At this juncture, we point out that equations (25) and (29) constitute the main analytic results of this work, and all which follows will be concerned with their validation and implications.

3.4 Ingredients for evaluating the covariance

To evaluate the covariance matrix we need to provide models for $n(M)$, $b(M)$ and the Fourier transform of the survey window function.

To compute $n(M)$ and $b(M)$ we employ the mass function and bias models presented in Sheth & Tormen (1999):

$$\frac{\mathrm{d}n}{\mathrm{d}\log M} = \frac{\bar{\rho}}{M}f_{\rm ST}(\nu)\frac{\mathrm{d}\log\nu}{\mathrm{d}\log M}\ ; \quad (30)$$

$$f_{\rm ST}(\nu) = A\sqrt{\frac{2q}{\pi}}\nu\left[1+(q\nu^2)^{-p}\right]\exp\left[-\frac{q\nu^2}{2}\right]\ ; \quad (31)$$

$$b_{\rm ST}(\nu) = 1+\frac{q\nu^2-1}{\delta_{\rm sc}}+\frac{2p/\delta_{\rm sc}}{1+(q\nu^2)^p}, \quad (32)$$

where $A=0.3222, q=0.707, p=0.3$. In the above we have introduced the peak-height $\nu(M)\equiv\delta_{\rm sc}/\sigma(M)$, where $\delta_{\rm sc}=1.686/D(z)$ is the spherical overdensity for collapse, and where $\sigma^2(M)$ is the variance of the linear density field extrapolated to $z=0$, smoothed with a spherical top-hat filter of radius R (see below for more details). This radius is defined so as to enclose a mass $M=4\pi\overline{\rho}R^3/3$, with $\overline{\rho}$ the mean matter density of the Universe at the present epoch.

For the survey window function we shall consider two simple examples. The first is a cubical top-hat, defined by:

$$W(\boldsymbol{x}|\boldsymbol{x}_j) = \begin{cases} 1/V_j\,, & x^l_j-L_{\rm box}/2\le x^l<x^l_j+L_{\rm box}/2 \\ 0, & \text{otherwise}\end{cases}, \quad (33)$$

where $l \in \{1, 2, 3\}$ denotes the Cartesian components of the vectors, j is the cell index, and $L_{\rm box}$ is the size of the cell of volume $V_j = L_{\rm box}^3$. The Fourier transform of this top-hat window function is:

$$W_j(\boldsymbol{k}) = \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}_j)\prod_{l=1}^{3} j_0\left(\frac{k_l L_{\rm box}}{2}\right), \tag{34}$$

where $j_0(y) \equiv \sin y/y$ is the zeroth order spherical Bessel function. The volume variance for this window function is:

$$\sigma^2(V_\mu) = \prod_{l=1}^{3}\left\{\int_{-\infty}^{\infty}\frac{\mathrm{d}k_l}{2\pi}\right\} P(k_1,k_2,k_3)\,|\widetilde{W}(\boldsymbol{k})|^2,$$
$$= 8\prod_{l=1}^{3}\left\{\int_{k_{\rm min}}^{k_{\rm max}}\frac{\mathrm{d}k_l}{2\pi}\right\} P(k_1,k_2,k_3)\,|\widetilde{W}(\boldsymbol{k})|^2, \tag{35}$$

where in the second equality we have used the isotropy of the power spectrum, e.g. $P(k_1, k_2, k_3) = P(-k_1, k_2, k_3)$. In equation (35) we use the following relation:

$$|\widetilde{W}(\boldsymbol{k})|^2 = \prod_{l=1}^{3} j_0^2\left(\frac{k_l L_{\rm box}}{2}\right). \tag{36}$$

The second window function is a spherical top-hat:

$$W_j(r) = \begin{cases} 3/(4\pi R^3), & |\boldsymbol{x}_j| < r < |\boldsymbol{x}_j| + R \\ 0, & \text{otherwise}, \end{cases}$$

where R is the radius of the spherical top-hat. The variance of the density field in this case has the familiar form:

$$\sigma^2(V_\mu) = \frac{1}{2\pi^2}\int_{k_{\rm min}}^{k_{\rm max}} \mathrm{d}k\, k^2 P(k)\widetilde{W}^2(kR) \tag{37}$$

for which the Fourier transform is given by:

$$\widetilde{W}(x) = \frac{3}{x^3}[\sin x - x\cos x]\,;\quad x \equiv kR. \tag{38}$$

On a technical note, we point out that for the k-space integrals given by equations (35) and (37), we have introduced lower and upper limits $k_{\rm min} > 0$ and $k_{\rm max}$, respectively. For a real survey, the upper limit is decided by the resolution of the instrument used. If the measurements are made from numerical simulations, which is the case with this work, the softening length of the simulations will dictate the largest frequency Fourier mode available: $k_{\rm max} = 2\pi/l_{\rm soft}$ and for our simulations $k_{\rm max} \sim 100\,h\,{\rm Mpc}^{-1}$. However, in practice the largest useful Fourier mode is much smaller, and occurs where the shot-noise correction to the power spectrum becomes comparable with the signal (Smith et al. 2003).

The lower limit $k_{\rm min}$ is a more complex issue. In the case of simulations, no modes with wavelength larger than the simulation box $L_{\rm sim}$ can contribute to the variance, which suggests the straightforward solution of adopting $k_{\rm min} = 2\pi/L_{\rm sim}$. Since we are attempting to confront the theory with the *reality* defined by simulations, we shall always assume this cut-off scale. However, for real surveys, the variance on a given scale will be affected by the existence of modes on scales larger than the size of the survey. We therefore recommend in this case $k_{\rm min} \to 0$, or at least the inverse horizon size at the redshift of the survey. For more discussion of the importance of $k_{\rm min}$ for the predictions of the variance, see discussion in Appendix A.

Note that in the above we shall relate the radius R of the spherical top-hat to that of the cubical top-hat function, through the relation $R = (3/4\pi)^{1/3} L_{\rm box}$. In other words the volumes of the spherical and cubical sample volumes are taken to be identical.

4 *N*-BODY SIMULATIONS

We study the covariance matrix with a suite of 40 large numerical simulations, executed on the zBOX-2 and zBOX-3 supercomputers at the Institute for Theoretical Physics, University of Zürich. For all realizations snapshots were output at: $z = \{5, 4, 3, 2, 1, 0.5, 0\}$. We shall refer to these simulations as the `zHORIZON` Simulations.

Each of the `zHORIZON` simulations was performed using the publicly available GADGET-2 code (Springel 2005), and followed the non-linear evolution under gravity of $N = 750^3$ equal-mass particles in a comoving cube of length $L_{\rm sim} = 1500\,h^{-1}\,{\rm Mpc}$. The cosmological model is similar to that determined by the *Wilkinson Microwave Anisotropy Probe* experiment (Komatsu et al. 2009). We refer to this cosmology as the fiducial model. The transfer function for the simulations was generated using the publicly available CMBFAST code (Seljak & Zaldarriaga 1996; Seljak et al. 2003), with high sampling of the spatial frequencies on large scales. Initial conditions were set at redshift $z = 50$ using the serial version of the publicly available 2LPT code (Scoccimarro 1998; Crocce, Pueblas &

Table 1. `zHORIZON` cosmological parameters. Columns are: density parameters for matter, dark energy and baryons; the equation of state parameter for the dark energy w; normalization and primordial spectral index of the power spectrum; dimensionless Hubble parameter.

Cosmological parameters	$\Omega_{\rm m}$	$\Omega_{\rm DE}$	$\Omega_{\rm b}$	w	σ_8	n	H_0 (km s^{-1} Mpc^{-1})
`zHORIZON-I`	0.25	0.75	0.04	−1	0.8	1.0	70.0
`zHORIZON-V1a/V1b`	0.25	0.75	0.04	−1	0.8	0.95/1.05	70.0
`zHORIZON-V2a/V2b`	0.25	0.75	0.04	−1	0.7/0.9	1.0	70.0
`zHORIZON-V3a/V3b`	0.2/0.3	0.7	0.04	−1	0.8	1.0	70.0
`zHORIZON-V4a/V4b`	0.25	0.8	0.04	−1.2/−0.8	0.8	1.0	70.0

Table 2. `zHORIZON` numerical parameters. Columns are: number of particles, box size, particle mass, force softening, number of realizations and total simulated volume.

Simulation parameters	$N_{\rm part}$	$L_{\rm sim}$ (Mpc h^{-1})	m_p(h^{-1} M$_\odot$)	$l_{\rm soft}$ (kpc h^{-1})	$N_{\rm ensemb}$	$V_{\rm tot}$ (h^{-3} Gpc^3)
`zHORIZON-I`	750^3	1500	5.55×10^{11}	60	40	135
`zHORIZON-V1, -V2, -V4`	750^3	1500	5.55×10^{11}	60	4	13.5
`zHORIZON-V3a`	750^3	1500	4.44×10^{11}	60	4	13.5
`zHORIZON-V3b`	750^3	1500	6.66×10^{11}	60	4	13.5

Scoccimarro 2006). Table 1 summarizes the cosmological parameters that we simulate and Table 2 summarizes the numerical parameters used.

In this paper we also study the Fisher matrix of cluster counts for which we use another series of simulations. Each of the new set is identical in every way to the fiducial model, except that we have varied one of the cosmological parameters by a small amount. For each new set we have generated four simulations, matching the random realization of the initial Gaussian field with the corresponding one from the fiducial model. The four parameter variations that we consider are $\{n \rightarrow \{0.95, 1.05\}$, $\sigma_8 \rightarrow \{0.7, 0.9\}$, $\Omega_{\rm m} \rightarrow \{0.2, 0.3\}$, $w \rightarrow \{-1.2, -0.8\}\}$, and we refer to each of the sets as `zHORIZON-V1a,b`,...,`zHORIZON-V4a,b`, respectively. Again, the full details are summarized in Tables 1 and 2.

Lastly, dark matter halo catalogues were generated for all snapshots of each simulation using the Friends-of-Friends (FoF) algorithm (Davis et al. 1985), with the standard linking-length parameter $b = 0.2$, where b is the fraction of the inter-particle spacing. For this we employed the fast parallel B-FOF code, kindly provided to us by V. Springel. The minimum number of particles for which an object is considered to be a bound halo was set at 20 particles. This gave a minimum host halo mass of $M \sim 10^{13}\,{\rm M}_\odot\,h^{-1}$.

5 RESULTS

In this section we confront the counts-in-cells theory with the results from N-body simulations.

5.1 Cell variance in simulations and theory

Since $\sigma^2(V_\mu)$ plays a vital role in determining the strength of any covariance in the mass function measurements, we shall make a detailed study of it, for both window functions discussed in Section 3 and considering volumes of varying size. We evaluate $\sigma^2(V_\mu)$ in two different ways, analytically and from N-body simulations. Furthermore, owing to concerns regarding the impact of non-linear bias and mass evolution, we also compute the matter–matter, halo–matter and halo–halo variance, which we denote as $\sigma^2_{\rm mm}(V_\mu)$, $\sigma^2_{\rm hm}(V_\mu)$ and $\sigma^2_{\rm hh}(V_\mu)$, respectively. Comparing these quantities will then make clear any departures from linearity.

Our analytical approach to determining the variances is based on standard quadrature routines to evaluate the theoretical expressions: for equation (35), we use the multi-dimensional Monte Carlo integration routine VEGAS; and for equation (37), we use the QROMB routine (for more details see Press et al. 1992). In evaluating these integrals we take the linear theory power spectrum matching our simulations, fully described in Section 4. Also, we take the largest mode in the simulation box to determine the lower limit of the k-integrals $k_{\rm min}$.

The second method is one of brute force: we measure $\sigma^2_{\rm mm}(V_\mu)$, $\sigma^2_{\rm hm}(V_\mu)$ and $\sigma^2_{\rm hh}(V_\mu)$ directly from the ensemble of simulations. Our estimator for the variances can be expressed as:

$$\hat{\sigma}^2_{\rm ab} \equiv \int \frac{{\rm d}^3\boldsymbol{k}}{(2\pi)^3} P_{ab}(k) W^2(kL_{\rm box}) \approx \frac{1}{V_\mu} \sum_{i,j,k=-N_{\rm g}/2+1}^{N_{\rm g}/2} \hat{P}_{\rm ab}(k_{ijk}) |W(k_{ijk}, L_{\rm box})|^2, \quad (39)$$

where the indices (i, j, k) label the Fourier mesh cell and k_{ijk} the magnitude of the wavenumber corresponding to that cell. The total number of grid cells considered is $N_{\rm g}^3$; also, a and b are $\in \{{\rm m}, {\rm h}\}$, and $\hat{P}_{ab}(\boldsymbol{k}_{ijk}) \equiv V_\mu \delta_a(\boldsymbol{k}_{ijk})^* \delta_b(\boldsymbol{k}_{ijk})$ are estimates of the various auto- and cross-power spectra. The window functions are as given in Section 3. The estimates of the variance also require a correction for shot-noise, which for the halo–halo variance we implement in the following way:

$$\hat{\sigma}^2_{\rm hh,c} = \hat{\sigma}^2_{\rm hh,d} - \frac{1}{N_{\rm h}} \sum_{i,j,k} |W(k_{ijk}, L_{\rm box})|^2, \quad (40)$$

where $N_{\rm h}$ is the number of haloes in the considered mass bin, and $\hat{\sigma}^2_{\rm hh,c}$ and $\hat{\sigma}^2_{\rm hh,d}$ are the variance of the continuous and discrete halo density fields, respectively. There is a similar shot-noise correction for the matter–matter variance; we assume that the halo-mass cross-variance requires no such correction. Note that the above method for estimating $\sigma(V_\mu)$ is not the conventional one, where one partitions the real space counts into cells and then computes the variance of that distribution. However, it should be entirely equivalent, but with the added advantages of being fast, since we are using a Fast Fourier Transform (FFT), and allowing for the computation of the variance in arbitrary cell structures.

Rather than testing all of the halo mass bins that we will employ later for the mass function covariance, we have chosen to show results for all the haloes in the simulation with $M > 10^{13}\,h^{-1}\,{\rm M}_\odot$. Fig. 1 presents our results for $\sigma_{\rm mm}(V_\mu)$, $\sigma_{\rm hm}(V_\mu)$ and $\sigma_{\rm hh}(V_\mu)$ as a function of the cubical window function size, $L_{\rm box}$; recall that for the spherical window we take the radius to be $R = (3/4\pi)^{1/3} L_{\rm box}$. In all three panels, the points represent results from the N-body simulations, whereas the lines denote the analytical integrals. The red full circles and solid lines are obtained by smoothing the density field with the cubical top-hat, while the blue empty circles and dashed lines denote smoothing with the spherical top-hat function. The simulation results represent the mean of the 40 realizations, with errors appropriate for a single run. The size of the simulation box ($L_{\rm sim} = 1500\,h^{-1}$ Mpc) is marked through a vertical black line on the horizontal axis. The effects of the shot-noise corrections on the estimates of $\hat{\sigma}^2_{\rm mm,c}$ and $\hat{\sigma}^2_{\rm hh,c}$ are too small to be noticed on this log–log plot.

As expected for a hierarchical mass distribution, in all cases the variance decreases steeply with the increasing box size. On comparing the results obtained from the simulations for the two window functions, we find very good agreement up until the size of the cubical region becomes similar to the size of the simulation cube. At this scale, the variance from the cubical window function displays a significant loss in signal. For scales larger than the simulation box, the smoothing result becomes somewhat meaningless and unstable due to the oscillatory nature of both window functions, which can be seen from the measurements.

Turning to the evaluation of the theoretical expressions for the variance, we see that in the case of the spherical top-hat there is excellent agreement between the simulations and the theory on small scales, $L_{\rm box} < 200\,h^{-1}$ Mpc. For $L_{\rm box} \geq 200\,h^{-1}$ Mpc, the linear expressions underestimate the measurements by $\approx$20 per cent or even more. However, on comparing the theoretical predictions for the cubical filter function with the measurements, we find a large discrepancy. We tested whether this was due to an error in the VEGAS evaluation of the integrals. An independent check with MATHEMATICA produced the same results.

After some investigations, we found that the discrepancy between the simulation and theory results was solely attributable to the difference between the discrete lattice structure of the Fourier space used in the simulations, and the continuum of Fourier modes used in the numerical integrals. A detailed discussion of this is presented in Appendix A1. In that section we also show that as the simulation

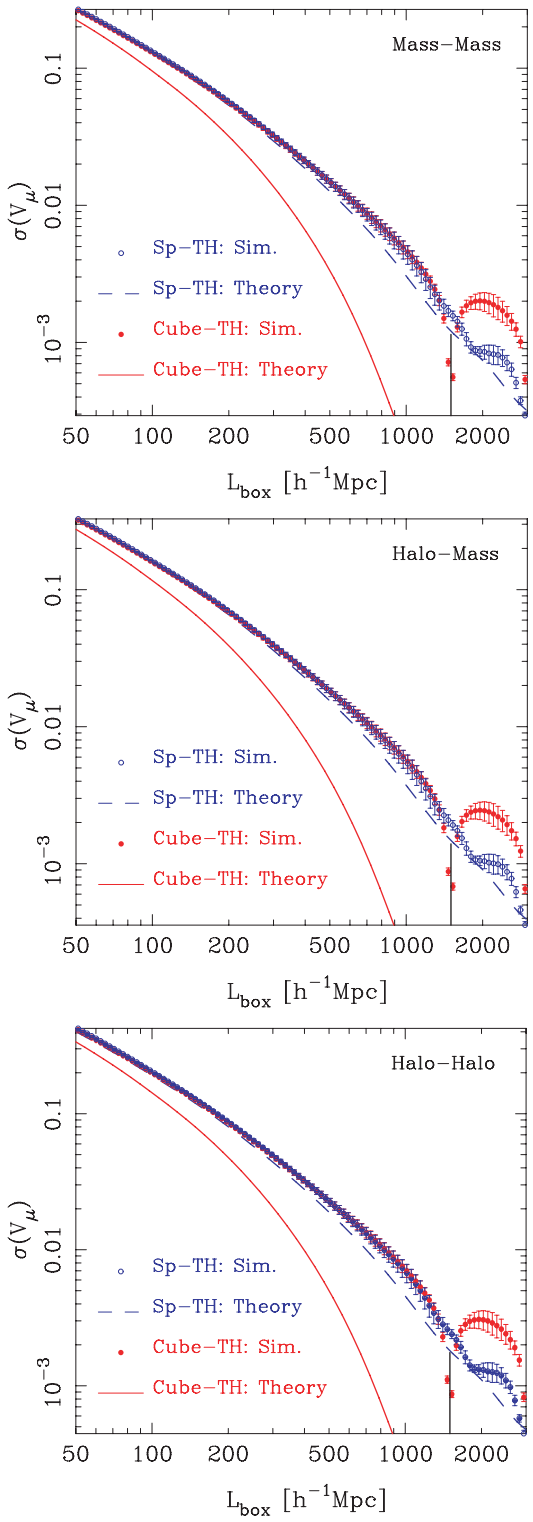

Figure 1. The rms density variance as a function of the sample volume size $L_{\rm box}$. From top to bottom, we show results for $\sigma_{\rm mm}(V_\mu)$, $\sigma_{\rm hm}(V_\mu)$, $\sigma_{\rm hh}(V_\mu)$, respectively. In each panel, blue empty and solid red circles denote measurements from the simulations, made using the spherical and cubical top-hat filter functions. The corresponding analytical predictions for the variance are denoted by the dashed blue and solid red lines, respectively. The size of the simulation box $1500\,h^{-1}$ Mpc is indicated by black vertical lines, and the measurements are an average of 40 simulations.

box size is increased, the theory and simulation results converge. Further, as is shown in Appendix A2 the theory predictions are sensitive to the lower limit $k_{\rm min}$. In applying this to the real Universe, we suggest letting $k_{\rm min} \to 0$.

5.2 Linearity of the bias

In linear theory, the relation between the variances plotted in Fig. 1 is given by:

$$\sigma^2_{\rm hh}(V_\mu) = \overline{b}\,\sigma^2_{\rm hm}(V_\mu) = \overline{b}^2\,\sigma^2_{\rm mm}(V_\mu). \tag{41}$$

$\overline{b}$ is the average linear bias from equation (7), estimated for a single mass bin containing all haloes larger than $10^{13}\,h^{-1}\,{\rm M}_\odot$. For the theoretical bias, we use the Sheth–Tormen model (Sheth & Tormen 1999, 2002), averaged over the same mass bin. All quantities are at redshift 0. Since the bias is >1, $\sigma_{\rm hh}(V_\mu)$ is slightly larger than $\sigma_{\rm hm}(V_\mu)$, which in turn is also slightly larger than $\sigma_{\rm mm}(V_\mu)$. At this level of detail the differences between the curves appear to be well related to each other as in equation (41).

To check this more accurately we next estimate the halo bias in the simulations and compare it directly with the theoretical predictions. In direct analogy with the Fourier-space bias estimates in Smith, Scoccimarro & Sheth (2007), we construct the following real-space bias estimates:

$$\hat{b}_{\rm hm} \equiv \frac{\sigma^2_{\rm hm}}{\sigma^2_{\rm mm}}; \quad \hat{b}_{\rm hh} \equiv \sqrt{\frac{\sigma^2_{\rm hh}}{\sigma^2_{\rm mm}}}, \tag{42}$$

where all quantities in the above depend on $L_{\rm box}$. Fig. 2 presents the comparison between the estimates of the linear bias from the simulations and the values obtained from the Sheth–Tormen formula. The top and bottom panels show the results for $b_{\rm hm}$ and $b_{\rm hh}$, respectively. Again, the solid red and empty blue circles denote the results from the cubical and spherical window functions, respectively. The Sheth–Tormen theory is represented by the thick green dashed line.

Considering $b_{\rm hm}$ (top panel), the first thing to remark is that the bias appears extremely flat over all of the scales probed – for the mean of the realizations the bias relation is linear to better than 1 per cent precision. Secondly, the peak-background split model of Sheth & Tormen predicts this value astonishingly well: $b = 1.498$.

Turning our attention to $b_{\rm hh}$ (lower panel), the raw simulation measurements (upper set of points) indicate that on scales $L_{\rm box} \geq 200\,h^{-1}$ Mpc, the bias displays a weak scale-dependence and is roughly $\sim$3 per cent higher than the Sheth–Tormen prediction. However, on smaller scales non-linear effects are apparent and the overall amplitude is steadily increasing with decreasing scale, being $\gtrsim$7 per cent higher than the Sheth–Tormen prediction for $L_{\rm box} = 50\,h^{-1}$ Mpc. The figure also shows the importance of correcting $\sigma^2_{\rm hh}$ for shot-noise when making estimates of the bias. The upper and lower set of points in this panel denote the uncorrected and corrected estimates, respectively. The shot-noise correction reduces the discrepancy between the simulations and linear theory to within $\sim$2 per cent for $L_{\rm box} \geq 200\,h^{-1}$ Mpc, however the non-linearity on smaller scales remains.

Both cubical and spherical window functions yield very similar results. In the rest of this work we shall employ the Sheth–Tormen bias, since on the scales of interest we have shown that it is at worst $<$5 per cent compared to the average bias of the haloes in our simulations.

Finally, we mention that for the analytical results in the following sections, we shall use: (i) the volume variance measured from the matter–matter power spectrum with a cubical window function, and not the analytical variance, given the discrepancy seen in Fig. 1. The

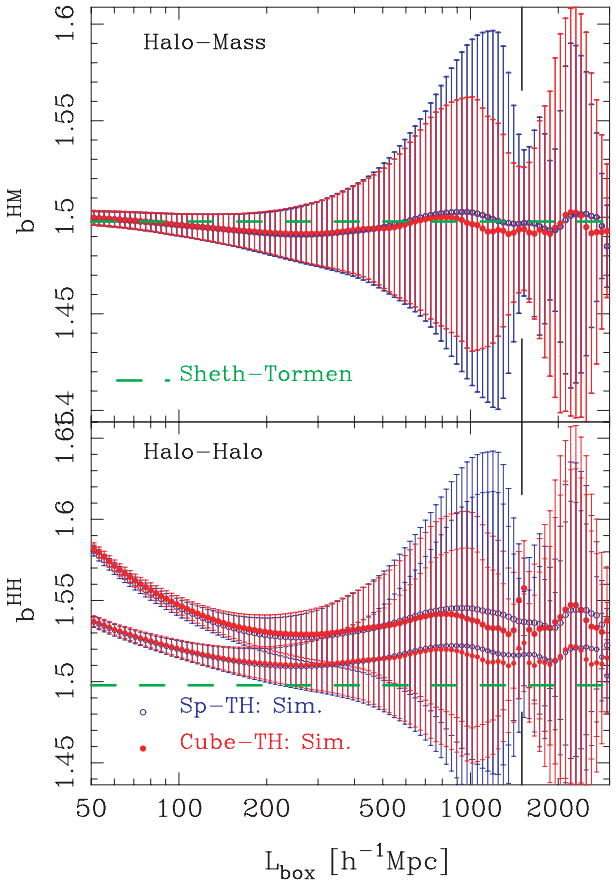

Figure 2. Comparison between the halo bias measured from the simulations and the Sheth–Tormen linear theory predictions as a function of the sample volume length. The symbols are as in the previous figure, and the theoretical prediction is represented by the dashed green line. The top panel shows the bias derived from the halo-matter variance, while the bottom panel shows the bias from the halo–halo variance. The lower panel also shows the importance of the shot-noise correction on the $b_{\rm hh}$ measurements: the upper and lower sets of points denote the halo–halo bias before and after the shot-noise correction, respectively.

cubical window function is a natural choice, since our simulations also have this geometry; (ii) the Sheth–Tormen bias; (iii) the Sheth–Tormen mass function.

5.3 An estimator for the mass function covariance

We estimate the mass function covariance matrix from the ensemble of 40 simulations of the fiducial cosmological model, described in Section 4. As we will show shortly, this number of realizations is insufficient for a reliable estimate of the covariance matrix. In order to overcome this problem, we have adopted the simple strategy of subdividing the volume associated with each realization into a set of smaller cubes. In particular, we divide each dimension of the original cube by 2, 3 and 4. Hence, each cube of $1500^3\,h^{-3}$ Mpc^3 is partitioned into 8, 27 and 64 subcubes with corresponding volumes of $750^3\,h^{-3}$ Mpc^3, $500^3\,h^{-3}$ Mpc^3 and $375^3\,h^{-3}$ Mpc^3, respectively. The 'subcubing' procedure thus provides us with 40, 320, 1080 and 2560 quasi-independent realizations. We note that this strategy was also adopted by Crocce et al. (2010), who used it to compute sample-variance error bars on the mass function in the MICE simulations. However, it has never been employed to compute the covariance matrix of counts.

One potential disadvantage of this approach is that the realizations thus obtained are not perfectly independent, since there will be modes with wavelength of the order of the initial box size $1500\,h^{-1}$ Mpc, which will potentially induce some covariance between the structures in each set of subcubes. However, as described in Appendix B, we have checked that this effect is of marginal importance. We shall therefore treat the measurements in each subcube as providing essentially independent information. Conversely, the subcubing approach should actually be thought of as the most relevant scenario, since in the real Universe there is no cut-off in the power spectrum on scales larger than the survey. As we demonstrated in Fig. 1, the cut-off scale in the simulations dramatically affects the behaviour of the density variance on the scales of the box. Hence, studying the mass function covariance using simulations that do not account for power on scales larger than the box modes may in fact lead to incorrect inferences about the real Universe.

Our estimator for the covariance matrix can be expressed as follows. Let $N_{\rm runs}$ be the total number of independent simulations in the fiducial suite, and $N_{\rm sc}$ the number of subcubes per simulation that we consider. For each subcube size, we estimate the average mass function as:

$$\hat{\bar{n}}(M_\alpha) = \frac{N_{\rm sc}}{V_{\rm sim}\Delta M_\alpha}\frac{1}{N_{\rm tot}}\sum_{i=1}^{N_{\rm tot}} N_{i,\alpha}\,, \tag{43}$$

where we defined $N_{\rm tot} = N_{\rm runs}{}^*N_{\rm sc}$ and $N_{i,\alpha}$ is the number of counts in the $i^{\rm th}$ subcube and mass bin α; $V_{\rm sim} = 1500^3\,h^{-3}$ Mpc^3, and $N_{\rm runs} = 40$. We estimate the mass function covariance between mass bins α and β:

$$\hat{\mathcal{M}}_{\alpha\beta} = \left(\frac{N_{\rm sc}}{V_{\rm sim}}\right)^2 \frac{1}{\Delta M_\alpha \Delta M_\beta}\frac{1}{N_{\rm tot}}\sum_{i,j=1}^{N_{\rm tot}} N_{i,\alpha}N_{j,\beta} - \hat{\bar{n}}(M_\alpha)\hat{\bar{n}}(M_\beta). \tag{44}$$

Note that in the above equation we subtract off the mean mass function averaged over all subcubes and all realizations in bins α and β. In order to check that the covariance matrix which we present below is not affected by our choice of the mean density of haloes, we recompute it using an alternative method: we determine the mean density for each realization and subtract it from the counts in the subcubes of that realization. This alternative is described in Appendix B. However, the results obtained from both methods are consistent.

The covariance matrices of the counts and the mass function are related through the equation:

$$\hat{C}_{\alpha\beta} = V_\mu^2 \Delta M_\alpha \Delta M_\beta \hat{\mathcal{M}}_{\alpha\beta}. \tag{45}$$

In the following sections, we present measurements made at $z = 0$. The mass function analysis is carried out for 12 logarithmically spaced bins, going from $(10^{13} < M\,[\,h^{-1}\,{\rm M}_\odot] < 10^{15})$. Finally, let us make the clarification that when we refer to 'halo mass', we mean the mass returned from the FoF algorithm.

5.4 Measurements: variance

Fig. 3 presents the fractional errors on the mass function, $\sigma[n(M)]/n(M)$, from both theory and simulations, for the subcube sizes mentioned in Section 5.3. In all panels the points denote the measurements from the simulations. The theoretical predictions of equation (25) are estimated for a single realization of given size $L_{\rm box}$, following the recipe at the end of Section 5.2. Then the variance is rescaled by $1/N_{\rm tot}$, so that the fractional errors in all four panels correspond to a total volume of 135 h^{-3} Gpc^3.

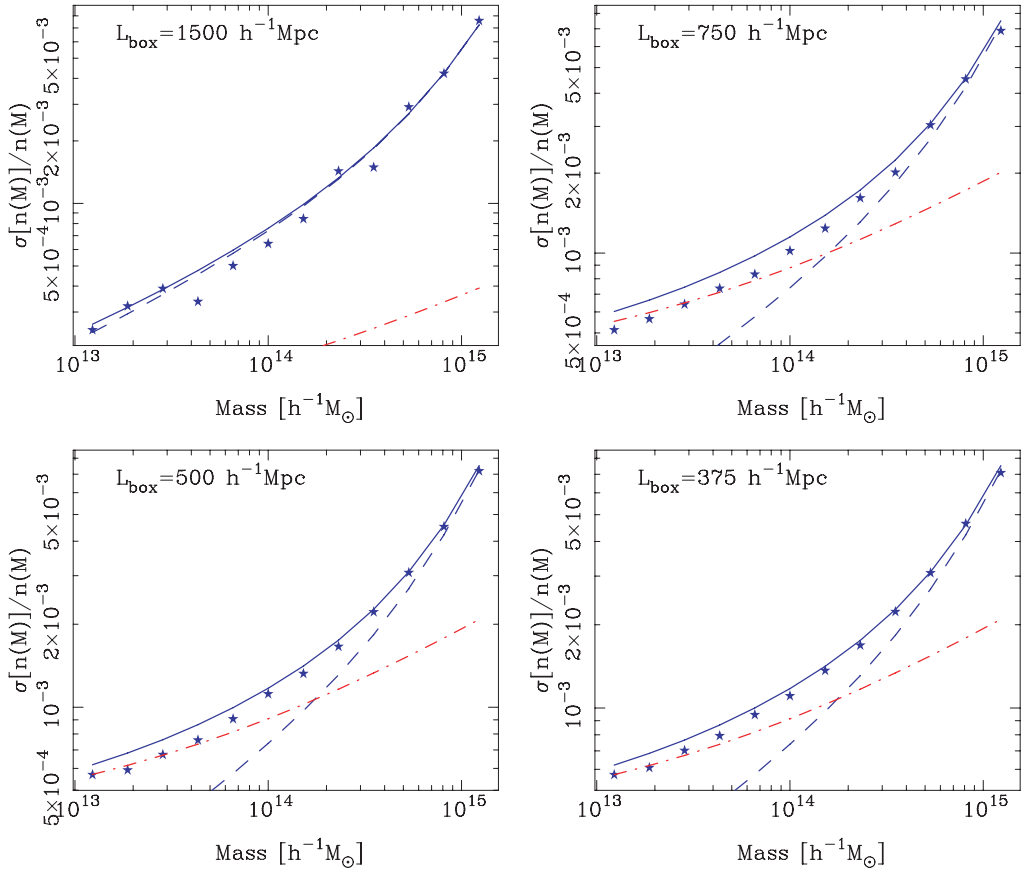

Figure 3. Comparison between the predicted and measured fractional error on the halo mass function as a function of halo mass. The four panels show the results obtained when the sample volume length is taken to be: $L_{\rm box} = \{1500\,, 750\,, 500\,, 375\,\}\,h^{-1}$ Mpc. In each plot, the dashed blue lines denote the fractional Poisson error; the red dot–dashed lines denote the pure sample variance error; and the solid lines represent the total. All errors have been rescaled to a total survey volume of $V = 135\,h^{-3}\,{\rm Gpc}^3$.

The agreement between the theory and the measurements is very good, with a slight difference at the low-mass end for the subcubes considered. This difference does not occur when the estimate is made using the full simulation boxes to estimate the variance (see the top left panel of Fig. 3). We also note that the Poisson model (dashed lines) agrees well with the simulations at the high-mass end. However, at lower masses, the variance becomes dominated by the sample variance, as given by the first term of equation (25). For a mass bin α the latter is simply:

$$\frac{\sigma[n(M_\alpha)]}{n(M_\alpha)} \approx \bar{b}_\alpha\,\sigma(V_\mu), \tag{46}$$

and this is denoted in Fig. 3 by the dot–dashed lines.

On comparing all four panels, we observe that with the exception of the first panel with $L_{\rm box} = 1500\,h^{-1}$ Mpc, the results are almost indistinguishable. This is quite interesting, since for these subcube volumes, Fig. 1 shows $\sigma(V_\mu)$ to be a decreasing function of $L_{\rm box}$. For a given mass bin we would expect the errors for the smaller subcube measurements to be significantly larger. This is indeed the case, but the fact that we use the variance on the mean, i.e. we divide by $\sqrt{N_{\rm tot}}$, leads to results that are very similar.

The slight difference between the measurements and the predictions is not easy to understand, since for the theoretical estimation we use $\sigma(V_\mu)$ measured from the simulations. This is done for all subcubes, so we do take into account that modes with wavelength larger than the subcube size may contribute to the covariance in the subcubes. The limit is set by the size of the original simulation box $L_{\rm sim} = 1500\,h^{-1}$ Mpc. However, the bias is the Sheth–Tormen prescription, which Fig. 2 shows to be slightly lower than the one measured from the halo–halo power spectrum. This effect might be more pronounced for the small-mass bins, but more work is needed here to arrive to a definitive conclusion, and we defer this to a future study.

Before moving on, we note that this startling agreement for the fractional errors on the mass function was noted before by Crocce et al. (2010). In that work the variance on a given subcube scale was computed theoretically using the linear theory variance in a spherical top-hat taken to have the same volume as the subcube (see the earlier discussion in Section 3). These authors pointed out that when using an ensemble of simulations with no subcubing the theory over-predicted the measurements. Here we have shown that there is no conflict between the theory and the measurements, if one uses the volume variance measured from the simulations.

5.5 Measurements: covariance

Fig. 4 presents the theoretical mass function correlation matrix from equation (29) versus the measured one. The left-hand panels show the predictions, obtained in the same way as in Fig. 3, and the right ones the measurements. From top to bottom, the following subcube sizes are considered: $L_{\rm box} = 375, 500, 750, 1500\ h^{-1}$ Mpc.

The figure reveals a remarkable agreement between measurements and theory: the trend observed in Fig. 3 is also present here,

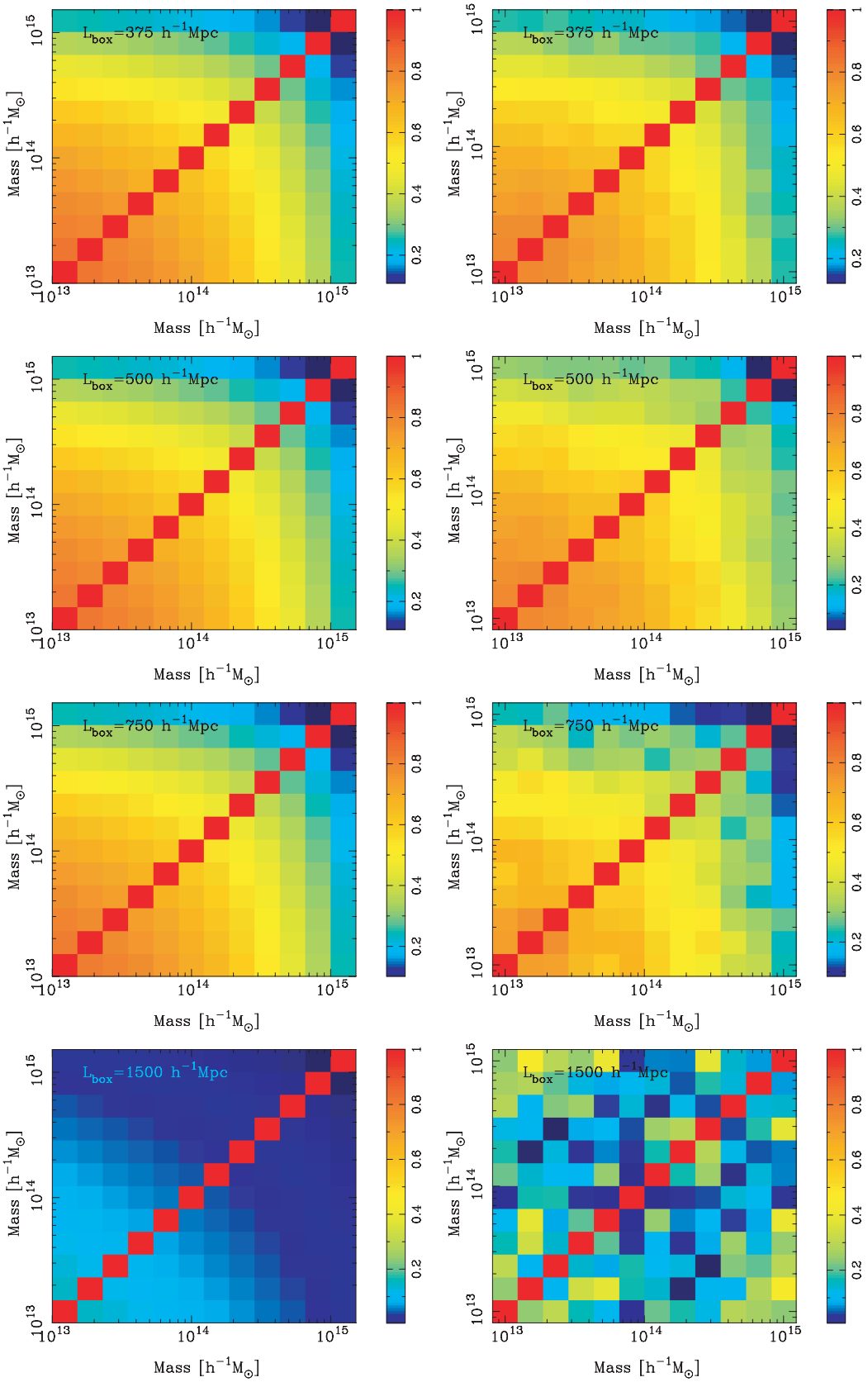

Figure 4. The correlation matrix of the cluster mass function $r_{\alpha\beta}$, i.e. equation (29). The left and right columns show the results from theory and simulations, respectively. From top to bottom, the size of the sample volume is given by: $L_{\rm box} = \{375, 500, 750, 1500\}\ h^{-1}$ Mpc. The theoretical predictions for the correlation matrix are generated using the estimate of $\sigma_{\rm mm}(V_\mu)$ measured directly from the simulations. Note that in the bottom-right panel we plot $|r_{ij}|$, so as to maintain the same heat-bar intensity scale as in the other plots.

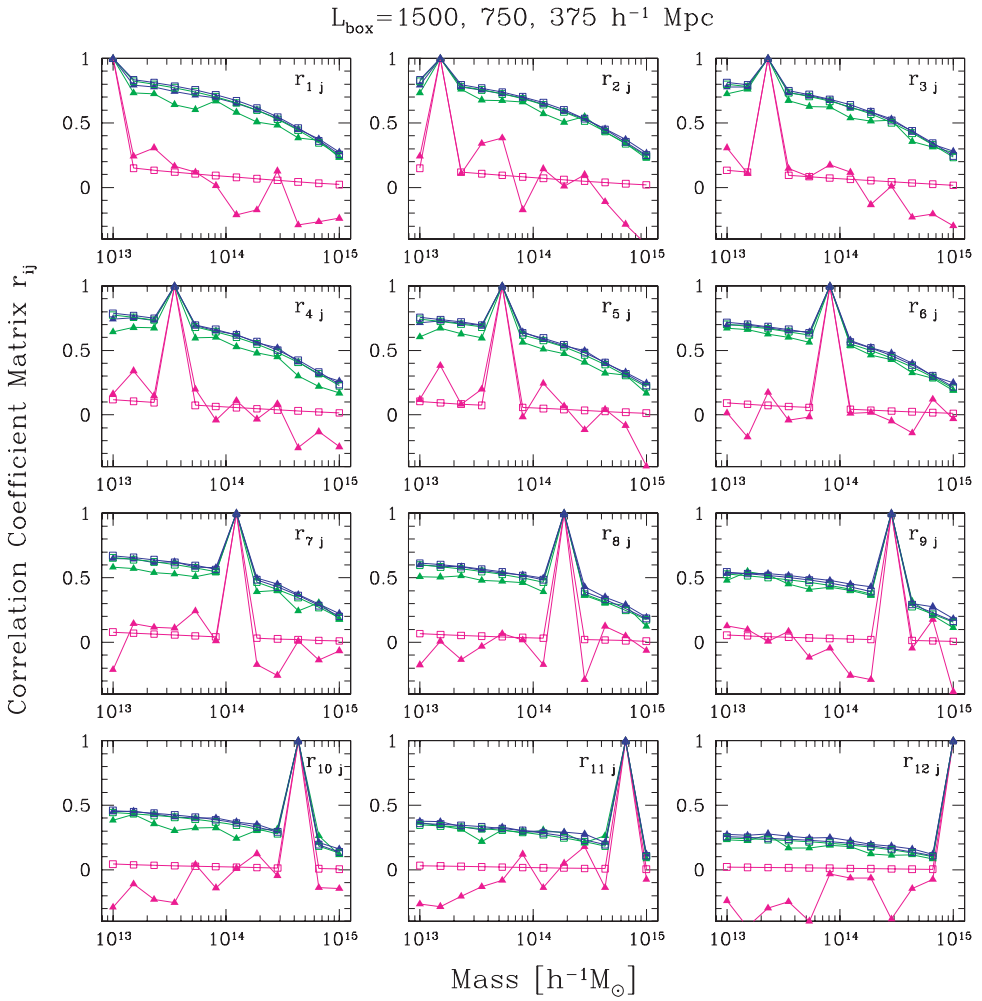

Figure 5. Rows of the cluster mass function correlation matrix $r_{\alpha\beta}$ as a function of the mass scale M_β, with M_α fixed. Each of the 12 panels shows the results for one of the 12 rows of $r_{\alpha\beta}$. In all panels, the theoretical predictions and the measurements from the simulations are denoted by the empty squares and solid triangle symbols, respectively. The magenta, green and blue colours represent the sample volume sizes $L_{\rm box} = \{1500, 750, 375\}\, h^{-1}$ Mpc, respectively.

with the predictions marginally larger than the measurements in some of the mass bins. We find that the measurements are strongly covariant: for clusters with $M \lesssim 3 \times 10^{14}\, h^{-1}\, {\rm M}_\odot$, the cross-correlation coefficient is $r \gtrsim 0.5$. Only for the highest-mass clusters does the covariance matrix become close to diagonal. The exception is for the ensemble of cubes with $L_{\rm box} = L_{\rm sim}$. In this case the realizations appear to be only weakly correlated with $r \gtrsim 0.1$, for $M \lesssim 3 \times 10^{13}\, h^{-1}\, {\rm M}_\odot$. However, as was discussed above and shown in Fig. 1, the behaviour of $\sigma(V_\mu)$ at the simulation box scale is not representative for the real Universe, owing to the absence of power on larger scales. Had we run a simulation of larger volume, then the volume variance on the scale $L_{\rm box} = 1500\, h^{-1}$ Mpc would have been significantly larger.

Fig. 5 presents the same information as Fig. 4, but in a more quantitative format. The plot has 12 panels, with each panel depicting a single row from the correlation matrix, i.e. $r_{ij}(M_i, M_j)$ versus M_j, with M_i fixed. In this plot the solid triangles denote the measured correlation coefficient, while the empty squares represent the theory predictions. For clarity, we show results only for the box sizes $1500, 750, 375\, h^{-1}$ Mpc, represented by the magenta, green and blue symbols, respectively. It is clear from this figure too that the theory predictions and the measurements are in remarkably good agreement. On comparing the correlation coefficient for different subcube sizes, we again note the similarity of these results, despite the variation in $\sigma(V_\mu)$: just as in Fig. 3, the covariance on the mean leads to the observed similarity. The exception is for the $L_{\rm box} = L_{\rm sim}$ cubes, and we offer the same explanation for this as noted above.

We conclude this section by stating that equation (25) gives a very reliable prediction for the mass function covariance, provided one employs the true variance within the volume.

6 COSMOLOGICAL INFORMATION FROM THE MASS FUNCTION

In this section we examine how the cosmological information content of the cluster mass function changes, when we exchange the standard Poisson assumption for the more complex likelihood models of equations (10) and (14).

6.1 Fisher information

In all cases we shall use the standard definition of the Fisher information (for an excellent review of Fisher matrix techniques in cosmology, see Heavens 2009):

$$F_{p_a p_b} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial p_a \partial p_b} \right\rangle , \qquad (47)$$

where p_a and p_b are elements of the cosmological model parameter set upon which the likelihood depends. From the Fisher matrix, one may obtain an estimate of the marginalized errors and covariances of the parameters:

$$\sigma^2_{p_a p_b} = [F^{-1}]_{p_a p_b}, \tag{48}$$

as well as the unmarginalized errors:

$$\sigma_{p_a} = [F_{p_a p_a}]^{-1/2}. \tag{49}$$

6.2 The Poisson Fisher matrix

In the case of Poisson errors, using equation (2) and (13) we write:

$$\begin{aligned}\ln \mathcal{L} &= \sum_{i,\alpha} \ln P(N_{i,\alpha}|\overline{m}_{i,\alpha}) \\ &= \sum_{i,\alpha} \left[-\overline{m}_{i,\alpha} + N_{i,\alpha} \ln \overline{m}_{i,\alpha} - \ln N_{i,\alpha}!\right].\end{aligned} \tag{50}$$

On partially differentiating the above expression with respect to parameters p_a and then p_b, and on performing the ensemble average, one finds:

$$F^{\rm Poisson}_{p_a p_b} = \sum_{i,\alpha} \frac{\partial \overline{m}_{i,\alpha}}{\partial p_a} \frac{\partial \overline{m}_{i,\alpha}}{\partial p_b} \frac{1}{\overline{m}_{i,\alpha}}. \tag{51}$$

6.3 The Gaussian Fisher matrix

As was shown earlier, in the case of the full likelihood model for the counts (cf. equation 10), we expect the Fisher matrix to be significantly modified from the Poisson case in the region of many counts per mass bin. In this limit, the likelihood is given by equation (14), and we have the standard result for the Fisher information for a Gaussian likelihood (Tegmark, Taylor & Heavens 1997):

$$\mathcal{F}^{\rm Gauss}_{p_a p_b} = \frac{1}{2} {\rm Tr} \left[\mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial p_a} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial p_b} \right] + \frac{\partial \overline{\boldsymbol{m}}}{\partial p_a}^{\rm T} \mathbf{S}^{-1} \frac{\partial \overline{\boldsymbol{m}}}{\partial p_b}. \tag{52}$$

6.4 The Gauss-Poisson Fisher matrix

LH04 developed an approximation for the Fisher matrix, which interpolates between the correct forms for the information in the limit of rare peaks and sample-variance-dominated counts. Their expression is:

$$\mathcal{F}^{\rm G+P}_{p_a p_b} \approx \frac{1}{2} {\rm Tr} \left[\mathbf{C}^{-1} \frac{\partial \mathbf{S}}{\partial p_a} \mathbf{C}^{-1} \frac{\partial \mathbf{S}}{\partial p_b} \right] + \frac{\partial \overline{\boldsymbol{m}}}{\partial p_a}^{\rm T} \mathbf{C}^{-1} \frac{\partial \overline{\boldsymbol{m}}}{\partial p_b}, \tag{53}$$

where $\mathbf{C} = \overline{\mathbf{M}} + \mathbf{S}$, and $\overline{\mathbf{M}}$ is a diagonal matrix with the elements $\overline{m}_{i,\alpha}$, as defined in Section 2.

6.5 Estimating Fisher matrices from simulations

In order to evaluate all of the expressions for the Fisher matrices presented in the previous sections, we require the knowledge of three quantities: the partial derivatives of the mean counts with respect to the parameters $\partial \overline{\boldsymbol{m}}/\partial p_a$; the inverse of the total covariance matrix $\mathbf{C}^{-1}$ and the derivative of the sample variance covariance matrix $\partial \mathbf{S}/\partial p_a$. In this section we shall use numerical simulations to directly evaluate all of these quantities.

We first measure the halo mass function for each of the variational cosmologies described in Section 4. With this information we are then able to numerically obtain the derivatives $\partial \overline{\boldsymbol{m}}/\partial p_a$ for the simulated parameters $p_a \in \{\Omega_{\rm m}, \sigma_8, n, w\}$. When computing the mass function derivatives, we reduce the effects of cosmic variance on the estimates, using the fact that the first four simulations of the fiducial cosmology have matched initial conditions with the four variational cosmologies simulations. Hence, our reduced-cosmic-variance estimator for the derivatives can be written as:

$$\frac{\partial \bar{n}_\alpha}{\partial p_b} = \frac{\bar{n}_\alpha}{N_{\rm var}} \sum_{r=1}^{N_{\rm var}} \frac{\partial \log \bar{n}^{(r)}_\alpha}{\partial p_b}, \tag{54}$$

where $\bar{n}_\alpha$ is the average mass function for mass bin α, estimated from all 40 independent realizations; $N_{\rm var} = 4$ is the number of the variational simulations; r denotes the simulation realization going from 1 to $N_{\rm var}$; $\bar{n}^{(r)}_\alpha$ is the mass function in the fiducial case, estimated for each of the four realizations that have matched initial conditions to the variational cosmologies realizations (for an explicit definition of this see equation (B1)). The logarithmic derivatives are estimated as:

$$\frac{\partial \log \bar{n}^{(r)}_\alpha}{\partial p_b} = \frac{\bar{n}^{(r)}_\alpha(p_b + \Delta_b) - \bar{n}^{(r)}_\alpha(p_b - \Delta_b)}{2\Delta_b \bar{n}^{(r)}_\alpha(p_b)}. \tag{55}$$

Note that since we estimate $\partial \overline{\boldsymbol{m}}/\partial p_a$ using double-sided derivatives, we may take larger step sizes in the parameters to compute the derivatives than would be allowed for single-sided derivatives (Eisenstein, Hu & Tegmark 1999). For the former case, the errors in the derivatives are of quadratic order in the step size: i.e. $\Delta[\partial \overline{\boldsymbol{m}}/\partial p_a] \approx (\Delta p_a)^2 \partial^3 \overline{\boldsymbol{m}}/\partial p_a^3/6$. Thus parameter step sizes of 20 and 10 per cent should correspond to relative errors of roughly 4 and 1 per cent in the derivatives, respectively. In actuality, the true accuracy of the derivatives also depends on the value of the third partial derivative.

In Fig. 6 we show simulation measurements of the average mass functions for the fiducial and variational cosmologies. This figure makes very clear not only the sensitivity of the mass function to the cosmological parameters considered, but also the halo mass range over which most of it occurs. Changes in $\Omega_{\rm m}$ and the slope of the primordial power spectrum n impact the mass function for the whole range of halo masses. The low-mass end is less sensitive to variations in σ_8, while the dark energy equation of state parameter w barely affects the mass function.

In the smaller panels of Fig. 6 we show the derivatives of the halo abundance, estimated using equation (55). The error bars are computed as errors on the mean of the $N_{\rm var} = 4$ realizations, as they are also for the mass functions in the larger panels. The $\Omega_{\rm m}$-derivative is almost constant and large for all bins, while the σ_8 one monotonically increases from 0 at the low-mass end to a large value at the high-mass end. The behaviour of the spectral-index-derivative is quite interesting, as it changes sign at $M \sim 3 \times 10^{14}\ h^{-1}\,{\rm M}_\odot$ and becomes negative at the high-mass end. Its overall variation is not as large as in the case of $\Omega_{\rm m}$ and σ_8, which will be better constrained by the halo abundance.

Another interesting finding of this exploration concerns the w-derivative, which should be 0 at redshift 0 according to linear theory and the Sheth–Tormen mass function. We find it to be small and positive, $\sim$0.05, for most of the mass range considered, and rising slightly to $\sim$0.1 at the low-mass end. The w-derivative does not vanish because in reality the mass function depends on the full non-linear growth history. This encompasses the growth of structure at all redshifts, and thus makes the present-day halo abundance

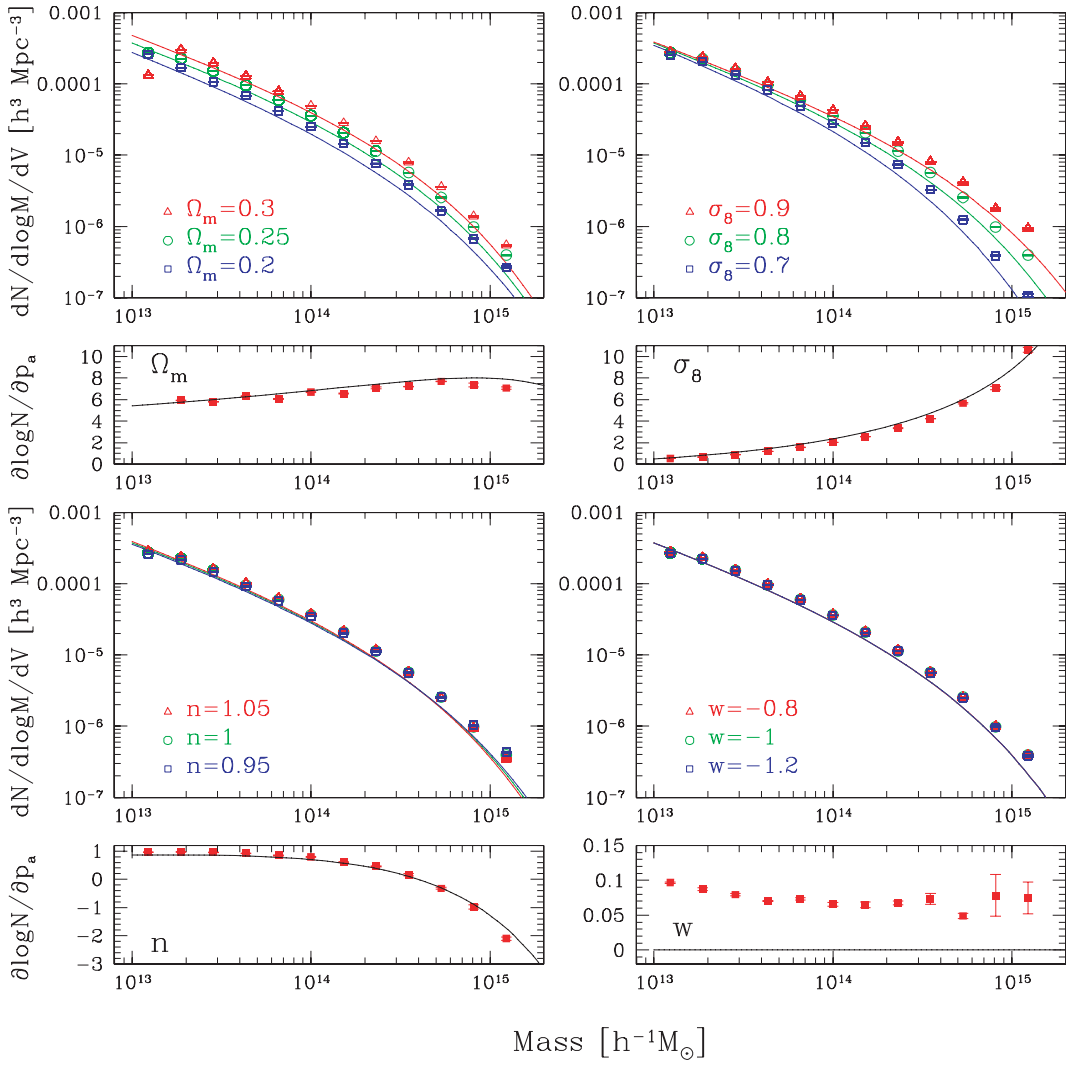

Figure 6. Top section of each panel: dependence of the $z = 0$ cluster mass function on cosmology, as a function of cluster mass. Symbols denote measurements from the simulations and lines depict the Sheth & Tormen (1999) mass function. The green colour represents the fiducial model, whereas the red/blue colours are for the plus/minus variations in the parameters. Bottom sections: logarithmic derivatives of the cluster number counts with respect to the considered parameters, as a function of cluster mass. Points with errors denote measurements from the simulations (cf. equation 55), the error bars being on the mean. Lines denote the Sheth–Tormen predictions.

sensitive to w. These results are consistent with the findings in earlier studies (Linder & Jenkins 2003; Jennings et al. 2010).

We next follow the recipe of Section 5.3 to estimate the covariance matrices in each of the variational cosmological models. From these estimates we are then able to form the partial derivatives of the covariance with respect to the cosmological parameters: $\partial \mathbf{C}/\partial p_a$. Again, as was done for $\partial \overline{\boldsymbol{m}}/\partial p_a$, we take advantage of the matched initial conditions to reduce the cosmic variance on the estimates of the partial derivatives of the covariance matrix.

6.6 Forecasted errors

Having obtained all of the necessary ingredients we are now in a position to evaluate the Fisher information directly from the simulations.

Fig. 7 shows the cumulative fractional Fisher errors, $\Delta p_a/p_a^{\rm fid}$, estimated using equation (49), as a function of the minimum cluster mass, and for the four cosmological parameters that we consider. The results obtained for the various subcube sizes are almost identical with the exception of the case where $L_{\rm box} = L_{\rm sim}$: as explained earlier, the underestimate of the variance on scales of the simulation makes the estimate of the mass function covariance, and hence the Fisher errors unrealistic. For brevity we shall present only the findings for $L = 375\,h^{-1}$ Mpc, which we consider very reliable.

For our fiducial survey, we adopt parameters relevant for future all-sky X-ray cluster surveys, such as eROSITA (Predehl et al 2010). This mission will be able to target intermediate-mass range clusters, and not just the most massive objects in the Universe as is the case for current and past surveys. We adopt a total survey volume of $V \sim 13.5\,h^{-3}\,{\rm Gpc}^3$, and we rescale our measured covariance matrices to this volume. For this comoving volume at $z = 0$, we find in the simulations approximately 4.5×10^6 haloes in the mass interval $[1, 5] \times 10^{13}\,h^{-1}\,{\rm M}_\odot$, 5.4×10^5 haloes in the interval $[0.5, 1] \times 10^{14}\,h^{-1}\,{\rm M}_\odot$, 2.3×10^5 haloes in the interval $[1, 6.5] \times 10^{14}\,h^{-1}\,{\rm M}_\odot$ and 8000 haloes with masses larger than the latter limit.

In Fig. 7 the solid green squares denote the results obtained for the Poisson Fisher matrix, as given by equation (51). The solid

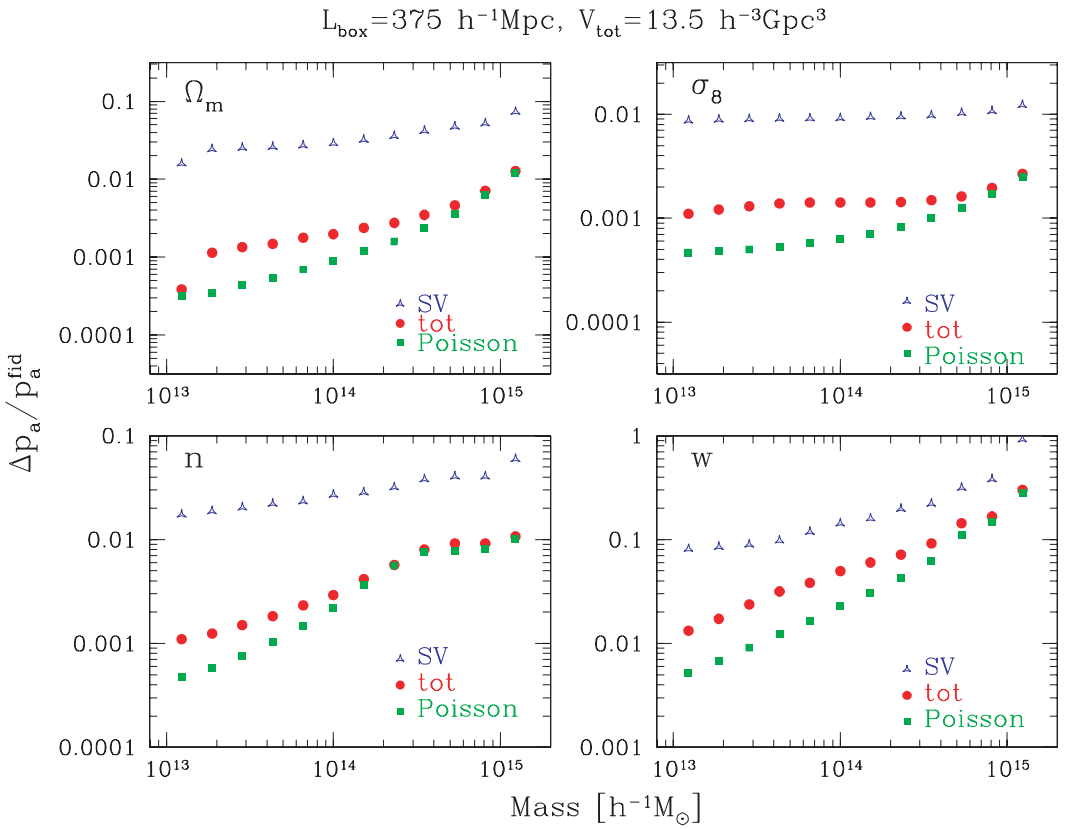

Figure 7. Fractional Fisher matrix errors on the cosmological parameters, as a function of the minimum cluster mass used. The four panels show the results for the cosmological parameters: $p_\alpha \in \{\Omega_m, \sigma_8, n, w\}$. In all panels, the symbols show the estimates made from the N-body simulations, with varying assumptions about the form of the cluster likelihood function. Solid green squares denote the Poisson errors obtained with equation (51); solid red circles denote the errors obtained from the second term of equation (53); blue triangular shaped symbols denote the errors derived from the trace-term in equation (53).

red circles denote the errors resulting from the second term of equation (53). The blue triangular-shaped symbols denote the errors obtained from only the trace-part of equation (53), where instead of $\mathbf{S}$, we have used the covariance matrix from our simulations $\hat{\mathbf{C}}$ (cf. equation 45). We do not expect that replacing $\hat{\mathbf{C}}$ with $\mathbf{S}$ will change our conclusions concerning the information carried by this term, except to possibly make the errors larger.

As expected, for all of the cosmological parameters considered, the fractional errors obtained from the Poisson approximation are smallest. Including the full covariance matrix, as in the second term of equation (53), reduces the amount of information, and this results in a significant increase in the fractional errors. For the case of $M_{\rm min} \sim 10^{13}\,h^{-1}\,{\rm M}_\odot$, the errors are roughly a factor of $\sim$3 larger when the full-covariance is used as opposed to the Poisson case. When $M_{\rm min} \sim 10^{14}\,h^{-1}\,{\rm M}_\odot$, the errors are only a factor of $\sim$2 worse. For the rarest objects, where the covariance becomes almost diagonal, the errors from the two methods are very similar. We find that the trace part of equation (53) contributes negligibly to the information, and if this term is taken separately, it yields errors that are roughly one order of magnitude larger than those from the second term.

Let us explore the consequences of this last result a little further. Consider the Fisher matrix given by equation (53), if the first term on the right-hand-side is negligible, then the information about each cosmological parameter enters the system only through the derivatives of the model mean with respect to the parameters. Since the model here is the mean counts, the bias provides no information. However, the amplitude of the elements of the information matrix can be modulated by the inverse covariance matrix. Owing to the fact that increasing the elements of the covariance matrix only leads to a smaller inverse covariance, we thus conclude that adding the variance from the bias can only ever decrease the Fisher information. However, as discussed in Lima & Hu (2005), the information content of the first term of equation (53) becomes of great importance in the presence of a scatter between the true and observed mass.

Note also that the cumulative dependence of the errors on the mass bins can partly be understood by examining the behaviour of the derivatives as shown in Fig. 6. The errors flatten out at those points in the mass range where the derivatives of the parameters are close to 0, as in the case of σ_8 at the low-mass end, or n at masses $\sim 3 \times 10^{14}\ h^{-1}\,{\rm M}_\odot$.

Finally, we emphasize that the forecasts that we make above are to illustrate the importance of going beyond the Poisson likelihood approximation and should not be taken as serious predictions for a potential survey. The cosmological dependence that we have considered here arises strictly from the mass function. In order to make a realistic forecast we would have to take into account a number of observational factors: realistic survey geometries; the evolution of the mass function with redshift; the evolution of the volume element with the cosmological model; and the evolution in the minimum detectable mass at each redshift; and a scatter in the relation between the observed mass proxy and the true cluster mass (see Marian & Bernstein 2006, for an example of forecasting weak lensing cluster counts).

7 SUMMARY AND CONCLUSIONS

In this paper, we have studied the covariance of the halo mass function, and the cosmological information content of such data. We adopted a two-line attack on these problems: the first line was theoretical and we developed an analytic model to explore these issues; the second was the use of a large ensemble of numerical simulations to measure directly all quantities of interest.

In Section 2, we summarized the counts-in-cells formalism (Peebles 1980; HK03), and developed it for application to deal with cluster counts in multiple mass bins. We described the Gauss–Poisson likelihood function for the counts in cells with multiple mass bins. The expression was analogous to that derived by LH04 for multiple cells and a single mass bin.

In Section 3, we used this framework to derive a formal expression for the covariance of the halo mass function and the cross-correlation coefficient, i.e. equations (25) and (29), respectively. We found that there were two terms contributing: a Poisson shot-noise term, which dominates in the limit of rare clusters, and a term associated with the sample variance, which is dominant for abundant clusters. This expression is analogous to the results of HK03 for multiple cells and a single mass bin. The expression was found to depend on three quantities: the cluster mass function; the cluster bias and the variance in the survey volume.

In Section 4, we presented the details of our large ensemble of numerical simulations: 40 simulations of a fiducial model and 32 simulations of modified cosmological models.

In Section 5, we made a rigorous comparison of the results from the theoretical modelling with those obtained directly from the numerical simulations. We measured the variance of matter and cluster fluctuations in cells of various sizes and found, for spherical and cubical top-hat cells, that the simulations and theory predictions were discrepant for large cell sizes. We showed that this was entirely attributable to the difference between the discrete lattice structure of the Fourier space in the simulations, and the continuum of Fourier modes in the theory integrals. The cubical and spherical top-hat simulation results were in good agreement, except on the largest scales where simulation box-scale effects were important.

We also measured the halo bias in cells of various sizes from the simulations. We found that the bias from the halo-mass cross-variance showed very little scale dependence over the range $L_{\rm box} = [50,\ 1500]\,h^{-1}\,{\rm Mpc}$, whereas that from the halo auto-variance showed significant scale dependence, before and after the shot-noise correction. We found that the Sheth & Tormen (1999) model was an excellent fit to the former and a reasonable fit to the latter.

We then measured the covariance of the mass function in the simulations. To increase the number of realizations, we used the strategy of subdividing each large simulation volume into a set of smaller subcubes. We found that the estimated covariances were in excellent agreement with the theoretical predictions. This was under the condition that we used the actual variance of mass fluctuations measured in the simulations.

In Section 6, we employed the Fisher matrix formalism to explore the information content of the cluster counts. Using the more realistic likelihood functions, we demonstrated numerically that the Poisson likelihood model only provides a reasonably accurate description of the data for clusters that are more massive than $M \gtrsim 5.0 \times 10^{14}\,h^{-1}\,{\rm M}_{\odot}$. Future surveys that aim to target cluster samples with masses $M \lesssim 5 \times 10^{14}\ h^{-1}\,{\rm M}_{\odot}$ must adopt more sophisticated likelihood analysis, such as discussed by LH04, Hu & Cohn (2006) and here in, which take into account the full covariance matrix of the counts. Otherwise, significant underestimates of the true errors will occur.

There are a number of possible future directions for the work that we have presented here. First, as pointed out by Lima & Hu (2005), one of the main uses of adopting the counts-in-cells approach is that it helps to lift the degeneracy between nuisance parameters, which are involved in calibrating the cluster masses, and the cosmological parameters. This occurs because the sample variance depends on the bias of the clusters, which has a different behaviour with cosmological parameters than the counts. Whilst we have shown explicitly that the terms in the Fisher matrix that depend on the derivatives of the covariance matrix, and hence derivatives of the bias, do not carry a great deal of cosmological information by themselves, it will be interesting to see whether for a more realistic scenario, where one must marginalize over these nuisance parameters, the self-calibration can be successfully performed to restore the lost information.

We also note that the counts-in-cells technique has been highlighted as a powerful means for constraining primordial non-Gaussianity (Oguri 2009; Cunha, Huterer & Doré 2010; Marian et al. 2011). It is of some importance to explore this approach using numerical simulations, since it is not clear whether the extension of the current formalism to such modified cosmological models works in practice.

ACKNOWLEDGMENTS

We thank Gary Bernstein and Ravi Sheth for comments on the draft and Raul Angulo, Martin Crocce, Peter Schneider, Uros Seljak and Yu-Ying Zhang for useful discussions. We thank V. Springel for making public GADGET-2 and for providing his B-FoF halo finder, and R. Scoccimarro for making public his 2LPT code. RES acknowledges support from a Marie Curie Reintegration Grant, an award for Experienced Researchers from the Alexander von Humboldt Foundation and partial support from the Swiss National Foundation under contract 200021-116696/1. LM is supported by the Deutsche Forschungsgemeinschaft through the grant MA4967/1-1.

REFERENCES

Abate A., Wittman D., Margoniner V. E., Bridle S. L., Gee P., Tyson J. A., Dell'Antonio I. P., 2009, ApJ, 702, 603
Albrecht A. et al,, 2006, preprint (arXiv e-prints)
Allen S. W., Schmidt R. W., Fabian A. C., Ebeling H., 2003, MNRAS, 342, 287
Borgani S. et al., 2001, ApJ, 561, 13
Cole S., Kaiser N., 1989, MNRAS, 237, 1127
Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369
Crocce M., Fosalba P., Castander F. J., Gaztañaga E., 2010, MNRAS, 403, 1353
Cunha C., Huterer D., Doré O., 2010, Phys. Rev. D, 82, 023004
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
DES 2005, preprint (arXiv e-prints)
Eisenstein D. J., Hu W., Tegmark M., 1999, ApJ, 518, 2
Gladders M. D., Yee H. K. C., Majumdar S., Barrientos L. F., Hoekstra H., Hall P. B., Infante L., 2007, ApJ, 655, 128
Heavens A., 2009, preprint (arXiv e-prints)
Henry J. P., 2004, ApJ, 609, 603
Hu W., Cohn J. D., 2006, Phys. Rev. D, 73, 067301
Hu W., Kravtsov A. V., 2003, ApJ, 584, 702 (HK03)
Israel H. et al., 2010, A&A, 520, A58
Jennings E., Baugh C. M., Angulo R. E., Pascoli S., 2010, MNRAS, 401, 2181
Komatsu E., The WMAP Team 2010, preprint (arXiv e-prints)

Komatsu E. et al., 2009, ApJS, 180, 330
Lima M., Hu W., 2004, Phys. Rev. D, 70, 043504 (LH04)
Lima M., Hu W., 2005, Phys. Rev. D, 72, 043006
Linder E. V., Jenkins A., 2003, MNRAS, 346, 573
LSST 2009, preprint (arXiv e-prints)
Manera M., Mota D. F., 2006, MNRAS, 371, 1373
Mantz A., Allen S. W., Ebeling H., Rapetti D., 2008, MNRAS, 387, 1179
Mantz A., Allen S. W., Rapetti D., Ebeling H., 2010, MNRAS, 406, 1759
Marian L., Bernstein G. M., 2006, Phys. Rev. D, 73, 123525
Marian L., Hilbert S., Smith R. E., Schneider P., Desjacques V., 2011, ApJ, 728, L13
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Muchovej S. et al., 2011, ApJ, 732, 28
Oguri M., 2009, Phys. Rev. Lett., 102, 211301
Okabe N., Zhang Y.-Y., Finoguenov A., Takada M., Smith G. P., Umetsu K., Futamase T., 2010, ApJ, 721, 875
Peebles P. J. E., 1980, The Large-scale Structure of the Universe. Research Supported by the National Science Foundation. Princeton Univ. Press, Princeton, NJ, p. 435
Planck Collaboration Ade P. A. R. et al.,, 2011, preprint (arXiv e-prints)
Predehl P. et al., 2010, in Arnaud M., Murray S. S., Takahashi T., eds, Proc. SPIE Vol. 7732, Space Telescopes and Instrumentation 2010: Ultraviolet to Gamma Ray. SPIE, Bellingham, 77320U
Press W. H., Schechter P., 1974, ApJ, 187, 425
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in FORTRAN. The Art of Scientific Computing, 2nd edn. Cambridge Univ. Press, Cambridge
Refregier A., Amara A., Kitching T. D., Rassat A., Scaramella R., Weller J., Euclid Imaging Consortium F. T., 2010, preprint (arXiv e-prints)
Reiprich T. H., Böhringer H., 2002, ApJ, 567, 716
Rozo E. et al., 2010, ApJ, 708, 645
Schirmer M., Erben T., Hetterscheidt M., Schneider P., 2007, A&A, 462, 875
Schuecker P., Böhringer H., Collins C. A., Guzzo L., 2003, A&A, 398, 867
Scoccimarro R., 1998, MNRAS, 299, 1097
Sehgal N., The ACT Collaboration, 2011, ApJ, 732, 44
Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437
Seljak U., Sugiyama N., White M., Zaldarriaga M., 2003, Phys. Rev. D, 68, 083507
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Sirko E., 2005, ApJ, 634, 728
Smith R. E. et al., 2003, MNRAS, 341, 1311
Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512
Springel V., 2005, MNRAS, 364, 1105
Sunyaev R. A., Zeldovich Y. B., 1972, Comments Astrophys. Space Phys., 4, 173
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Valageas P., Clerc N., Pacaud F., Pierre M., 2011, preprint (arXiv e-prints)
Vanderlinde K., The SPT Collaboration, 2010, ApJ, 722, 1180
Vikhlinin A. et al., 2009, ApJ, 692, 1060
Wang L., Steinhardt P. J., 1998, ApJ, 508, 483
Zhang Y.-Y., Finoguenov A., Böhringer H., Kneib J.-P., Smith G. P., Czoske O., Soucail G., 2007, A&A, 467, 437
Zhang Y.-Y., Finoguenov A., Böhringer H., Kneib J.-P., Smith G. P., Kneissl R., Okabe N., Dahle H., 2008, A&A, 482, 451

APPENDIX A: THE VOLUME VARIANCE

In this section, we investigate the impact of systematic effects on the volume variance, which arise due to the anisotropic lattice in Fourier space and also low- and high-k truncation of the matter power spectrum.

A1 Fourier lattice versus a continuum of modes

As was described in Section 3.4 the matter variance in the volume is a key quantity for correctly evaluating the covariance matrix of the cluster counts. Also, as was shown in Section 5.1, there is a discrepancy between the theoretical predictions and measurements in simulations obtained for $\sigma(V_\mu)$. We now investigate the origin of these discrepancies.

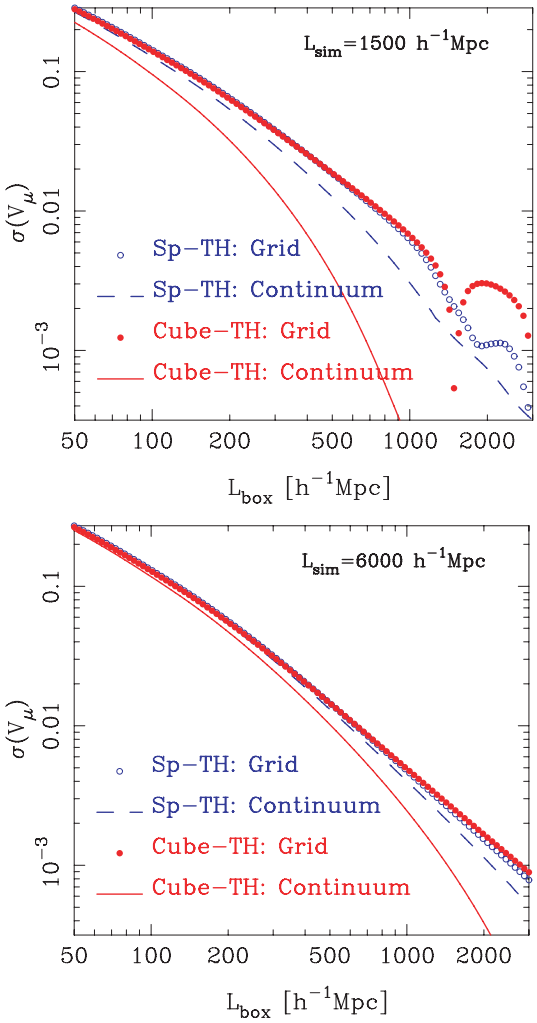

Figure A1. The rms density variance $\sigma_{\rm mm}(V_\mu)$ as a function of the sample volume size $L_{\rm box}$. In each panel, blue empty and solid red circles denote theoretical predictions made on the Fourier space lattice using the spherical and cubical top-hat filter functions, respectively. The predictions made using a continuum of Fourier space modes are denoted by the dashed blue and solid red lines, respectively. The top panel compares the results when the simulation box size is taken to be $L_{\rm sim} = 1500\,h^{-1}$ Mpc. The bottom panel shows the same but for the case where $L_{\rm sim} = 6000\,h^{-1}$ Mpc.

We start by examining the importance of the discrete cubical Fourier lattice, which is used in the estimates from the simulations, and the continuum of Fourier modes, which is used to evaluate the theory. We start by generating a Fourier lattice as in the simulations, where each lattice point is spaced from the next one, along each dimension, by $k_{\rm f} = 2\pi/L_{\rm sim}$. Then at each lattice point, we compute the magnitude of the k-vector and evaluate the linear theory power spectrum at that point. $\sigma(V_\mu)$ is then obtained as described in equation (39), by summing up the grid of power spectra values multiplied by the square of the appropriate window function. The top panel of Fig. A1 shows the results of this exercise for both the spherical and cubical window functions. We also compare this to the results obtained from the theory, assuming a continuum of modes. The results that we find for the theory evaluated on the cubical mesh

are in remarkably good agreement with the measurements from the simulations presented in Fig. 1.

To be sure that the discrepancy is due to the lattice, we should expect that as the simulation box size becomes significantly larger, the results for the lattice should approach those of the continuum. We test this by regenerating the Fourier lattice, but this time taking $L_{\rm sim} = 6000\,h^{-1}$ Mpc, and keeping the maximum Fourier mode the same as before. The bottom panel of Fig. A1 shows that the results are now in much better agreement.

Thus, we are led to conclude that in matching the results from the simulations we must be mindful of taking into account the anisotropic lattice structure of the Fourier space to obtain accurate comparisons between the theory and the simulations. This then further justifies our choice of using the $\sigma(V_\mu)$ measured in the simulations to make the predictions for the covariance of the counts.

Finally, these results also act as a cautionary tale: when interpreting the results of numerical simulations on very large scales, we should take more care in assigning the power to the lattice cells in the initial conditions. We should use methods that suppress this discretization. For instance, it would seem more sensible to compute the power averaged over a lattice cell and not simply the power at the lattice cell point. Also including the missing zero modes may be a more realistic strategy (see e.g. Sirko 2005).

A2 Cut-off scales

In Fig. 1 we evaluated the integrals in equations (35) and (37), keeping the upper and lower bounds fixed at the values $k_{\rm min} = 2\pi/L_{\rm sim} = 0.004\,h\,{\rm Mpc}^{-1}$ and $k_{\rm max} = 2\pi/l_{\rm soft} = 100\,h\,{\rm Mpc}^{-1}$. This was done for a fair comparison with our simulations, which do not have modes larger than the simulation box $L_{\rm sim} = 1500\,h^{-1}$ Mpc, nor structures smaller than the softening scale, $l_{\rm soft} = 0.06\,h^{-1}$ Mpc. We now present a short discussion of how the mass-fluctuations-variance $\sigma(V_\mu)$ depends on the cell volume and the cut-off scales in the power spectrum.

For the large cell sizes that we are interested in, i.e. $L_{\rm box} > 50\,h^{-1}$ Mpc, we find no dependence of $\sigma(V_\mu)$ on $k_{\rm max}$, for the range of values $k_{\rm max} = [1, 100]\,h\,{\rm Mpc}^{-1}$.

For the lower cut-off scale $k_{\rm min}$, the situation appears to be more complex. In Fig. A2, we show the result of computing the mass-fluctuations-variance averaged in cubical and spherical top-hat volumes, as a function of the cubical cell volume [recall that we take the radius of the spherical top-hat cell to be $R = (3/4\pi)^{1/3} L_{\rm box}$]. In the plot we consider the values of $\sigma(V_\mu)$ for three different simulation sizes: $L_{\rm sim} = \{750, 1500, 3000\,h^{-1}\,{\rm Mpc}\}$. These box sizes correspond to the: solid red, long-dashed green and dot–dashed magenta lines, respectively. The thicker/thinner lines in the plot depict the spherical/cubical top-hat smoothing.

For the case of the spherical top-hat filter, we find that the variance for $k_{\rm min} = 2\pi/750\,h\,{\rm Mpc}^{-1} = 0.008\,h\,{\rm Mpc}^{-1}$ is roughly a factor of $\sim$2 times smaller than the variance obtained when $k_{\rm min} = 2\pi/3000 = 0.002\,h\,{\rm Mpc}^{-1}$. However, for the case of the cubical top-hat window function, we find that the difference in the variance for these same two values of $k_{\rm min}$ is more than an order of magnitude.

In Fig. A2, the thick dotted blue curve presents predictions for $\sigma(V_\mu)$ with the spherical window function. Hence, the Covariance matrix can be written as $k_{\rm min} = 2\pi/L$. The thick and thin dashed blue lines show the same, but for the case of the cubical filter function. For this case, the thin line is obtained when the linear theory matter power spectrum is used, and the thicker line shows the results obtained when the non-linear power spectrum from `halofit` (Smith et al. 2003) is employed. The differences are very small. Thus, using the linear theory power spectrum for the mass variance is quite reasonable on these scales. The main point of this last example is to show that for large cell sizes, the results for $\sigma(V_\mu)$ are very sensitive to the presence/absence of power on very large scales.

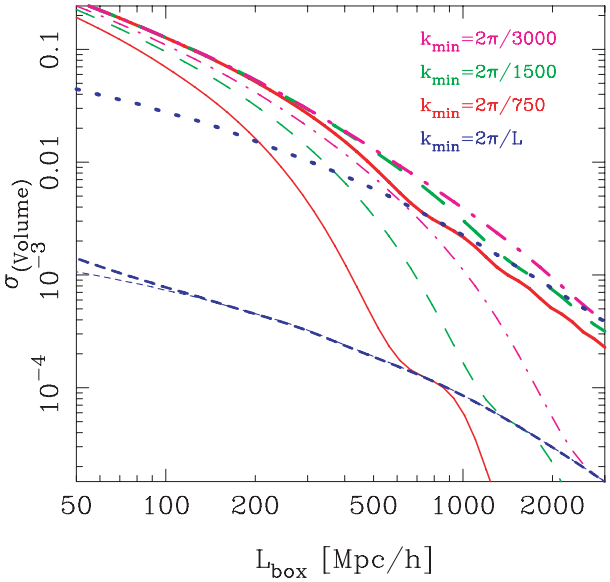

Figure A2. Dependence of the rms density variance on the lower limit $k_{\rm min}$ of the k-space integrals, as a function of the sample volume size $L_{\rm box}$. The thin solid red, dashed green and dot–dashed magenta lines denote results obtained for a cubical top-hat window function, where $k_{\rm min} = 2\pi/L_{\rm sim}$ with $L_{\rm sim} = \{750, 1500, 3000\}\,h^{-1}$ Mpc, respectively. The thick solid, dashed and dot–dashed lines represent the same, but for the case where the filter function is a spherical top-hat. The thick and thin blue dotted lines denote the same as above except this time the lower limit of the k-space integrals is given by $k_{\rm min} = 2\pi/L_{\rm box}$.

APPENDIX B: CONVERGENCE OF THE COVARIANCE MATRIX

B1 Covariances from individual simulations

Here we consider an alternate approach to estimating the covariance of the cluster counts. We are concerned that if there is a significant variance of the cluster counts on the scales of the simulation cube, then by computing the covariance around the mean cluster mass function averaged over all simulations, we are overestimating the covariance. To answer this question, we adopt the strategy of using the subcubes in a single simulation to make an estimate of the covariance, and finally we then average these estimates over all the simulations.

For each simulation run we therefore have:

$$\overline{n}_r(M_\alpha) = \frac{N_{\rm sc}}{V_{\rm sim}\Delta M_\alpha}\,\frac{1}{N_{\rm sc}}\sum_{i=1}^{N_{\rm sc}} N_{i,\alpha}^{(r)}\,, \tag{B1}$$

where $N_{i,\alpha}^{(r)}$ is the number of counts in the $r^{\rm th}$ run, $i^{\rm th}$ subcube and $\alpha^{\rm th}$ mass bin. The covariance for each run is:

$$\mathcal{M}_{\alpha\beta}^{(r)} = \left(\frac{N_{\rm sc}}{V_{\rm sim}}\right)^2 \frac{1}{\Delta M_\alpha \Delta M_\beta}\,\frac{1}{N_{\rm sc}}\sum_{i,j=1}^{N_{\rm sc}} N_{i,\alpha}^{(r)} N_{j,\beta}^{(r)} - \overline{n}_r(M_\alpha)\overline{n}_r(M_\beta)\,, \tag{B2}$$

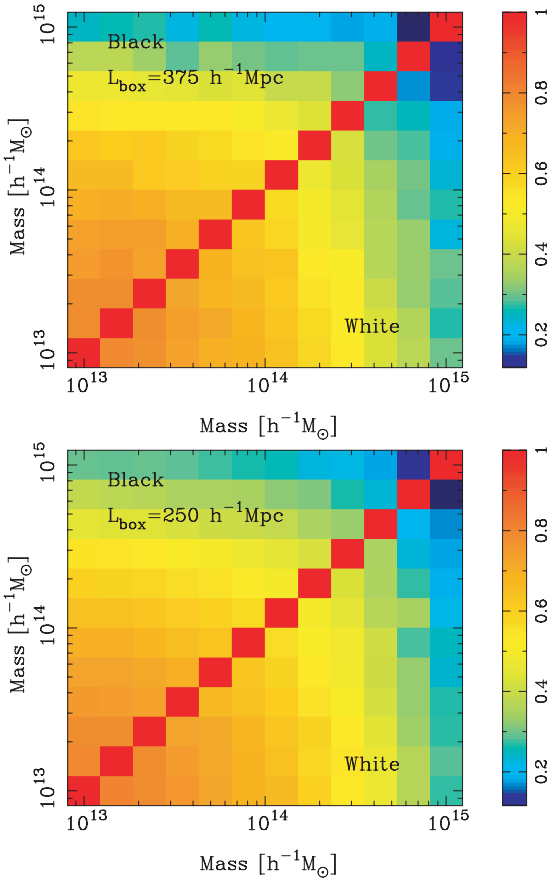

Figure B1. The chessboard test: we compare the mass function correlation matrix measured from 'white' and 'black' subcubes (see the Appendix text). The results are very similar for both subcube sizes considered, $375^3\ h^{-3}\,\mathrm{Mpc}^3$ and $250^3\ h^{-3}\,\mathrm{Mpc}^3$.

and the average covariance:

$$\mathcal{M}_{\alpha\beta} = \frac{1}{N_{\rm runs}} \sum_{r=1}^{N_{\rm runs}} \mathcal{M}^{(r)}_{\alpha\beta}. \tag{B3}$$

We have checked that using equations (B1) and (B2) does not change the measured mass function covariance in any significant way. We therefore conclude that the method of estimating the covariance described in Section 5.3 is not biased by the estimates of the mean density.

B2 The chessboard test

When dividing a big simulation box into smaller subcubes, the largest wavelength modes may affect the observables measured in the smaller cubes. In the case of clusters, some of the subcubes may have very different mean counts than others, and in general, the smaller the subboxes, the larger the expected covariance between them. This is also true for real surveys, which measure observables in a finite volume of the Universe: some of these observables are impacted by modes larger than the size of the survey.

In order to check the validity of our approach, we measure the covariance of the mass function using subcubes that are not adjacent, and should therefore be less covariant. We shall refer to this as the 'chessboard test', as its 2D analogue would be similar to using only the white or the black squares of a chessboard to compute the mass function covariance. This test has the limitation that large mode correlations can span more than just two subcubes, particularly if the latter are small. Nevertheless, if we find the covariance measured from the 'white' subcubes different from that obtained from the 'black' ones and also different from the all-subcubes-covariance, then our box-division method is flawed.

We perform this test for the conservative values $n = 4, 6$, i.e. we consider 4^3 and 6^3 subcubes, with volumes $375^3\ h^{-3}\,\mathrm{Mpc}^3$ and $250^3\ h^{-3}\,\mathrm{Mpc}^3$, respectively. The result is shown in Fig. B1. There is no major difference between the 'white', 'black' and total mass function correlation matrix (cf. Fig. 4). We conclude that modes with wavelengths smaller than the size of the subcubes considered here are not explicitly responsible for generating the mass function covariance.

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