

Towards accurate field-level inference of massive cosmic structures

Stephen Stopyra¹  ¹★ Hiranya V. Peiris,^{1,2} Andrew Pontzen,² Jens Jasche¹ and Guilhem Lavaux³

¹The Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, AlbaNova, Stockholm SE-106 91, Sweden

²Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

³Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France

Accepted 2023 October 9. Received 2023 September 22; in original form 2023 May 3

ABSTRACT

We investigate the accuracy requirements for field-level inference of cluster and void masses using data from galaxy surveys. We introduce a two-step framework that takes advantage of the fact that cluster masses are determined by flows on larger scales than the clusters themselves. First, we determine the integration accuracy required to perform field-level inference of cosmic initial conditions on these large scales by fitting to late-time galaxy counts using the Bayesian Origin Reconstruction from Galaxies (BORG) algorithm. A 20-step COLA integrator is able to accurately describe the density field surrounding the most massive clusters in the local super-volume ($< 135 h^{-1}$ Mpc), but does not by itself lead to converged virial mass estimates. Therefore, we carry out ‘posterior resimulations’, using full N -body dynamics while sampling from the inferred initial conditions, and thereby obtain estimates of masses for nearby massive clusters. We show that these are in broad agreement with existing estimates, and find that mass functions in the local super-volume are compatible with Λ CDM.

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

1 INTRODUCTION

Traditionally, cosmological constraints have relied on observables constructed from summary statistics of the density field, such as the power spectrum. By contrast, the technique of *field-level inference* – in which the full posterior distribution of the density field is sampled – potentially allows one to access additional information contained e.g. in the phases of the density field. Examples of the application of field-level inference include a determination of the local matter density from the 2M++ galaxy catalogue with the Bayesian Origin Reconstruction from Galaxies (BORG) algorithm (Jasche & Lavaux 2019), the inference of the COSMOS initial density field using Lyman- α data by Horowitz et al. (2019), Porqueres et al. (2019b), and Ata et al. (2022), and the use of effective field theory by Babić, Schmidt & Tucci (2022) to infer the density field on the baryon acoustic oscillation scale from halo catalogues. Field-level inference has also been demonstrated to outperform two-point statistics for weak lensing (Porqueres et al. 2022; Porqueres et al. 2023), and its robustness has previously been investigated in the context of effective-field-theory (EFT) likelihoods (Kostić et al. 2022; Nguyen et al. 2021). However, accurate field-level inference on scales that are even mildly non-linear at late times is challenging for a number of reasons. The dynamic range of gravitational collapse and the astrophysical complexity of galaxy biasing are key issues that must be addressed in order to accurately infer a density field from observational data.

As an example of the potential power of field-level inference, cluster masses have long been envisioned as a probe of the cosmo-

logical parameters (Bocquet et al. 2016; Costanzi et al. 2019; Pratt et al. 2019) and of beyond- Λ -Cold-Dark-Matter (Λ CDM) physics, such as primordial non-Gaussianity (Sartoris et al. 2010; LoVerde & Smith 2011; Stopyra et al. 2021b) or modified gravity (Mak et al. 2012; Ilić, Sakr & Blanchard 2019). However, clusters are structures on scales of order Mpc, which is very small compared with the overall volume in which the inference takes place. Therefore, directly inferring cluster masses using field-level inference within a traditional Bayesian sampling framework would require spatial (and time-stepping) resolution that remains, for now, computationally intractable.

The mass of clusters is none the less expected to be physically dictated by large-scale flows (Bertschinger 1985; Lucie-Smith, Peiris & Pontzen 2019; Lucie-Smith, Adhikari & Wechsler 2022). Density and velocity information at such scales can be accurately inferred with currently available approximate dynamical models such as FastPM (Feng et al. 2016) or CO-moving Lagrangian Acceleration (COLA Tassev, Zaldarriaga & Eisenstein 2013). This opens up the possibility of using field-level inference to accurately infer the relevant initial conditions on larger scales. One may then resimulate, with high time and spatial resolution a number, of samples from the posterior on the initial density field. In this way, a posterior on the mass of a cluster as implied by the combination of larger-scale information and the gravity solver can be determined. We refer to this technique as posterior resimulation because it takes samples from the posterior distribution of initial conditions, and evolves each to redshift $z = 0$ with a higher-accuracy gravity solver.

Posterior resimulation is similar in spirit to the local Universe simulations of Heß, Kitaura & Gottlöber (2013). These belong in the broader landscape of local Universe simulations, which has

* E-mail: stephen.stopyra@fysik.su.se

its genesis in Peebles (1989), who first studied initial conditions for simulations, which closely resemble the local Universe. Local Universe simulations are performed with a variety of techniques. For instance, the CLUES collaboration (Gottlöber, Hoffman & Yepes 2010; Sorce et al. 2014, 2016) use Hoffman–Ribak (Hoffman & Ribak 1991) and the Reverse Zel’dovich Approximation (Doumler et al. 2013) to incorporate a number of velocity constraints. A Hoffman–Ribak approach is also used by Lavaux (2010). The SIBELIUS simulations (McAlpine et al. 2022; Sawala et al. 2022) use field-level inference to set a large-scale environment, and subsequently introduce additional small-scale power, which is necessary to reproduce local-group structures. A similar approach is taken with the HESTIA hydrodynamical simulation suite (Libeskind et al. 2020), in which unconstrained small-scale modes are randomly seeded, and an ensemble of simulations with regions resembling the Local Group is obtained by selecting initial conditions that satisfy a set of criteria on the positions of cluster such as Virgo, and the locations/masses of Local Group galaxies.

In contrast to most local Universe simulations, *posterior resimulation* uses initial conditions drawn directly from the posterior distribution, and therefore accurately projects statistical uncertainties into the evolved universe, allowing us to explicitly assess the significance of inferred structures. This approach therefore offers a new probe of cluster masses and of other cosmological structure formation observables that are determined by information that resides at larger scales in the initial conditions, and is therefore strongly constrained by the combination of the gravitational collapse process and the large-scale environment. As such, posterior resimulation opens up new avenues for cosmological tests, e.g. one may compare the inferred cluster masses with independent estimates – from Sunyaev–Zel’dovich (SZ), X-rays or lensing for example – and for modelling or testing models for galaxy intrinsic alignments, which are believed to be sensitive to large-scale tidal fields (Codis, Pichon & Pogosyan 2015).

In this study, we use posterior resimulation of initial conditions obtained using field-level inference with BORG to estimate the masses of nearby galaxy clusters. We investigate how the accuracy of the gravity solver used for field-level inference affects cluster mass estimation using posterior resimulation. Informed by these results, we select an improved gravity solver and perform a new inference of the initial conditions, which achieves higher accuracy compared with those obtained by Jasche & Lavaux (2019). The latter initial conditions have been used by Desmond et al. (2022) to initialize simulations, and the resulting void properties have been studied. However, we demonstrate that the improved accuracy obtained with our choice of gravity solver is vital to reliably estimate both cluster and void masses. We leave a full discussion of void properties, such as their density profiles to future work.

The structure of the paper is as follows. In Section 2, we outline the methods used for field-level inference, posterior resimulation, and the validation of our results. In Section 3, we determine the accuracy of the gravity solver needed for the study and show the results for cluster and void mass functions. We then produce estimates of local massive cluster masses and validate our results via comparison with existing mass estimates, as well as through internal consistency checks. We discuss the cosmological implications of the results in Section 4, including an estimate of the underdensity of the local super-volume, and conclude in Section 5.

2 METHODS

In this work, we will focus on the problem of estimating the masses of galaxy clusters. We will use field-level inference to obtain the

distribution of initial conditions compatible with a galaxy catalogue, then resimulate these with greater accuracy to obtain the cluster masses themselves. The first step involves inferring the initial density field in the Lagrangian patch surrounding the galaxy clusters of interest. In particular, in order to obtain converged mass estimates via posterior resimulation in the second step, the large-scale flows in the vicinity of the cluster must be accurately inferred. This places strong requirements for the accuracy of the gravity solver used within the forward model in the field-level inference step, which we quantify in this study.

This section is structured as follows: in Section 2.1, we describe the field-level inference framework we use, including the forward model and the data set. In Section 2.2, we discuss how posterior resimulation is used to estimate cluster masses from the field-level inference. In Section 2.3, we outline the techniques we use for validating our results.

2.1 Field-level inference with BORG

In this work, we use the BORG (Jasche & Wandelt 2013) algorithm to perform field-level inference of galaxy cluster masses conditioned on the 2M++ galaxy catalogue (Lavaux & Hudson 2011). BORG uses a Hamiltonian Markov Chain Monte Carlo (MCMC) algorithm (Duane et al. 1987; Neal 1993) to sample the posterior distribution of possible initial density fields, δ_i^{IC} , assuming a Λ CDM Gaussian prior and conditioned on the observed galaxy counts, N_i , in a set of voxels (labelled by i). This posterior is given schematically by

$$P(\delta^{\text{IC}}|N) = \frac{P(\delta^{\text{IC}})P(N|G[\delta^{\text{IC}}])}{P(N)}, \quad (1)$$

where $P(\delta^{\text{IC}})$ is the Λ CDM prior on the initial conditions, and $P(N|G[\delta^{\text{IC}}])$ represents the likelihood of observing a given galaxy distribution given a specific set of the initial conditions. This is dependent on a *gravity solver*, $G[\delta^{\text{IC}}]$, which describes the gravitational evolution that maps initial densities onto the final density field at redshift $z = 0$. Additionally, a bias model is required to map the final density field into galaxy counts; while this is a crucial part of the inference, we do not represent this process in the schematic equation (1) above, since we will not assess the accuracy of bias modelling in this paper. We fix the bias model to that adopted by Jasche & Lavaux (2019), briefly outlined in Section 2.1.3, so that we can focus on the impact of the gravity solver choice on the inference accuracy.

Relative to previous work presented in Jasche & Lavaux (2019), we run a new MCMC inference with an improved gravity solver (see below) and the updated likelihood introduced by Porqueres et al. (2019a), which is further described in Appendix A.

2.1.1 The 2M++ galaxy catalogue

The data used in the BORG inference in this work is identical to that used by Jasche & Lavaux (2019), known as the 2M++ galaxy catalogue (Lavaux & Hudson 2011). This consists of targets drawn from the 2-Micron All-Sky Survey extended source catalogue (2MASS-XSC; Huchra et al. 2012), with spectroscopic redshifts from the 2MASS Redshift Survey (2MRS; Huchra et al. 2012). It additionally uses data from the 6-Degree Field galaxy redshift survey (6dFGRS; Jones et al. 2006), and the Sloan Digital Sky Survey Data Release 7 (SDSS; Abazajian et al. 2009).

Following Jasche & Lavaux (2019), we allow the BORG algorithm to infer galaxy bias parameters separately in several apparent and

absolute magnitude bins as follows. First, the galaxies are split into two apparent magnitude bins: $K \leq 11.5$ (reaching a distance of $\simeq 200 h^{-1}$ Mpc) and $11.5 < K \leq 12.5$ (reaching a distance of $\simeq 350 h^{-1}$ Mpc). The vast majority of the $\simeq 70\,000$ galaxies are therefore contained within the $677.7 h^{-1}$ Mpc simulation box. The latter does not contain any galaxies from 2MRS, due to incompleteness at fainter magnitudes.

These two apparent magnitude bins are further subdivided into eight absolute magnitude (denoted M_K) bins between $-25 \leq M_K \leq -21$, giving a total of 16 ‘catalogues’ in which the bias model parameters are determined independently. The forward model assumes that the dark-matter density field is common to all 16 catalogues.

2.1.2 Gravity solver

The choice of gravity solver within the forward model is crucial to the goal of estimating the cluster masses. It is important to note that clusters themselves will not be fully resolved within the field-level inference. However, the much larger Lagrangian volume in the initial conditions can be resolved, so once the linear field on large scales is accurately inferred, the flows into a cluster are determined. This is what enables the posterior resimulation to map the initial field onto accurate cluster masses. Hence, if the gravity solver does not reconstruct these large-scale flows accurately, the inferred initial conditions, and any derived quantities, such as galaxy cluster masses, will be biased. For example, if the gravity solver underestimates the non-linear growth in high-density environments, then the initial conditions will be driven towards artificially higher densities to match the galaxy catalogue. This would then result in the cluster mass being overestimated when these initial conditions are resimulated with an N -body code.

Since the initial conditions are inferred via sampling, the gravity-solver must also be computationally efficient, in order to obtain a converged Markov chain in a reasonable time. Accuracy and speed must therefore be carefully traded off against each other to satisfy the accuracy requirements of the problem under consideration.

In this work, we first used samples from the Jasche & Lavaux (2019) MCMC chain to test two gravity solvers: 1024³ particle mesh (PM; Eastwood & Hockney 1974; Klypin & Shandarin 1983; Klypin & Holtzman 1997), COmoving Lagrangian Acceleration (COLA; Tassev, Zaldarriaga & Eisenstein 2013), comparing them against the adaptive-time-stepping N -body code GADGET2 (Springel 2005). We consider the effect of spacing the time-steps linearly with scale factor, and logarithmically. We first investigated the optimal time-stepping procedure (see Section 3.1 for more details), and selected a 20-step COLA gravity solver (COLA20) with time-steps spaced linearly in scale factor to be used for our new MCMC inference. We compare our results to the 10-step particle mesh method (PM10) used to perform field-level inference using the same catalogue by Jasche & Lavaux (2019).

We use the solvers on a set of six initial conditions drawn from the Jasche & Lavaux (2019) MCMC chain to simulate the evolution of 512³ particles between $z = 69$ and 0 in a $677.7 h^{-1}$ Mpc box, for a final spatial resolution of $0.66 h^{-1}$ Mpc. Initial conditions are inferred on a 256³ grid, which are oversampled to produce the 512³ particles used by the gravity solvers. We compare different choices for the accuracy of the gravity solver in Section 3.

Because we perform the above gravity solver tests with initial conditions drawn from the Jasche & Lavaux (2019) MCMC chain, which was generated with the PM10 method, the masses shown in

Figs 1 and 2 are not expected to be consistent with the more accurate masses that result from our new MCMC inference below. However, the convergence properties of the gravity solver are not altered by this change in the initial conditions.

2.1.3 Galaxy bias model

The relationship between the *final* density field and the galaxy distribution, N_i , is described by a *galaxy bias model*. Specifically, the inference uses the Neyrinck et al. (2014) bias model, where the observed galaxy count N_i in voxel i is assumed to follow a Poisson distribution with mean number of galaxies λ_i . In the Neyrinck et al. (2014) model, this mean count is related to the final density contrast in that voxel, δ_i , by

$$\lambda_i(\delta_i, \bar{N}, \beta, \rho_g, \epsilon_g) = S_i A_{\alpha(i)} \bar{N} (1 + \delta_i)^\beta \exp(-\rho_g (1 + \delta_i)^{-\epsilon_g}). \quad (2)$$

The prefactor S_i accounts for the selection function and survey mask in voxel i , constructed following the procedure of Jasche & Lavaux (2019). The four parameters of the bias model (\bar{N} , β , ρ_g , ϵ_g) can be inferred jointly with the density field. In practice, we infer only β , ρ_g , and ϵ_g since we use the likelihood presented in Porqueres et al. (2019a), which is insensitive to \bar{N} (see Appendix A). Different parameters are inferred for each of 16 galaxy catalogues (each with a different absolute and apparent magnitude range) in the 2M++ data set.

The amplitude $A_{\alpha(i)}$ will be of particular interest in the present work. Its purpose is to account for possible unknown multiplicative, spatially-varying systematics, which may differ between each of the 16 catalogues. The $\alpha(i)$ subscript refers to the definition of the amplitudes over a separate healpix (Gorski et al. 2005) pixelization of the sky with $n_{\text{side}} = 4$, with each healpix pixel (healpixel) split into 10 radial bins of width $60 h^{-1}$ Mpc creating a set of 1920 regions labelled by α . Each cubic voxel, i , is uniquely found in a specific region, $\alpha(i)$, which is shared by a number of voxels (see Appendix A for further details). The values of A_α are marginalized over a Jeffreys prior in the likelihood, though it is necessary to reconstruct their posterior in order to perform the posterior predictive tests, which we outline in Section 2.3. As a reminder that A_α can differ between catalogues, later we will refer to these values as A_α^c .

2.1.4 Redshift space distortions

Redshift space distortions are treated as in Jasche & Lavaux (2019); the initial density field is first evolved to $z = 0$ using the gravity solver, which produces particles with known position and velocity. Then, the positions and velocities are combined to produce (physical) redshift space positions for each particle,

$$\mathbf{s} = \left(1 + \frac{a}{H(a)} \frac{\mathbf{v} \cdot \mathbf{r}}{|\mathbf{r}|^2} \right) \mathbf{r}, \quad (3)$$

where $H(a)$ is the Hubble rate as a function of scale factor a , $\mathbf{r} = a\mathbf{x}$ is the position in physical units, \mathbf{x} the comoving position, and $\mathbf{v} = d\mathbf{x}/dt$ is the comoving velocity. The density field in redshift space is then computed using the Cloud-in-Cell (CIC) approach on a 256³ grid, with a spatial resolution of $2.65 h^{-1}$ Mpc. By applying the bias model in equation (2) to the redshift space density field on this grid, we can compute mean galaxy counts for each voxel, which are then compared to the 2M++ galaxies on the same grid using the likelihood.

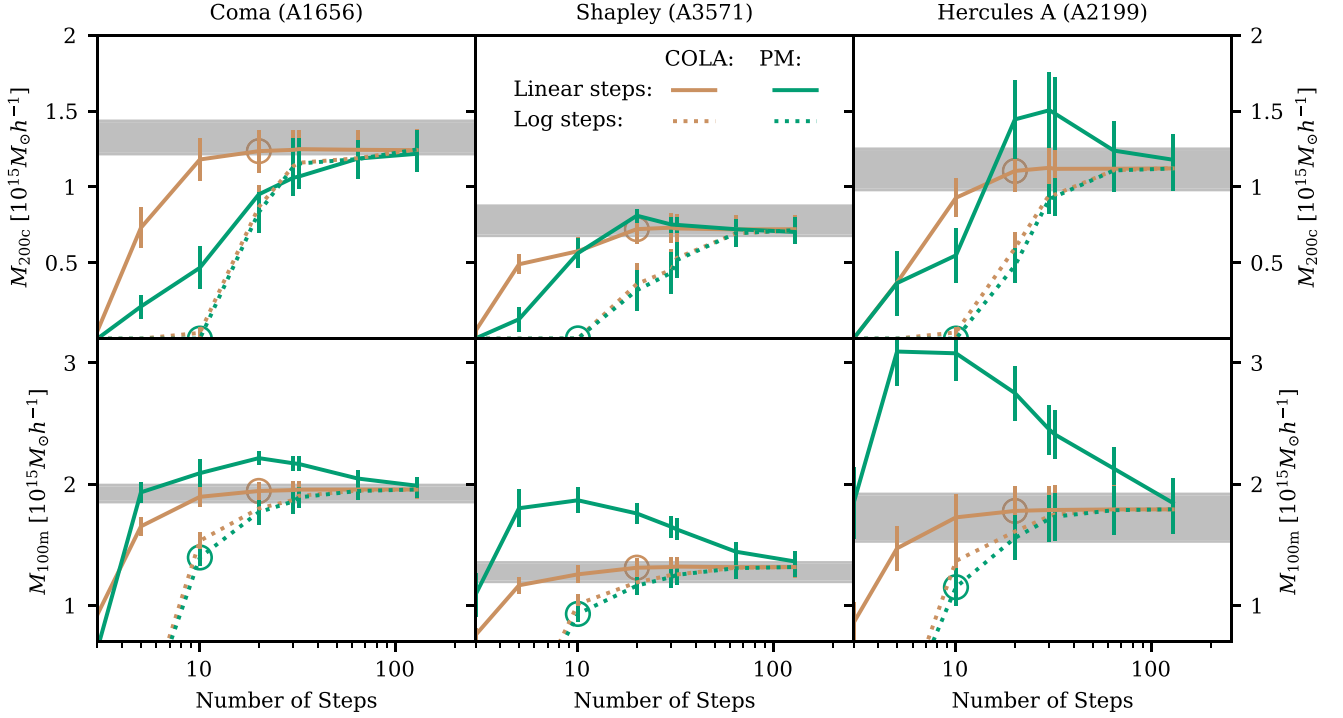


Figure 1. Convergence of the M_{200c} (upper row) and M_{100m} (lower row) masses of three $\sim 10^{15} M_{\odot} h^{-1}$ clusters with COLA (brown lines) and PM (green lines) gravity solvers, starting from identical initial conditions from the Jasche & Lavaux (2019) BORG inference. Solid lines indicate time-steps spaced linearly in scale factor, and dotted lines indicate time-steps spaced logarithmically in scale factor. Circles highlight the results for COLA20 (brown), which is used for the chain computed in this work, and PM10 (green), which was used for the chain generated by Jasche & Lavaux (2019). Error bars denote the standard deviation of the mean over all contributing resimulated MCMC samples. The mass of the clusters is compared to that obtained using posterior resimulation (grey region, showing the mean and the standard deviation of the mean mass over six samples from the Markov chain). The PM10 model fails to accurately describe the mass enclosed even within the large scales measured by M_{100m} , but COLA accurately describes the mass at this scale.

2.1.5 Cosmological parameters and numerical convergence

For our new MCMC inference with the COLA solver, we assumed the Planck 2018 (Planck Collaboration et al. 2020) cosmological parameters with lensing and baryon acoustic oscillations: $\Omega_m = 0.3111$, $\sigma_8 = 0.8102$, $H_0 = 67.66 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $n_s = 0.9665$, $\Omega_b = 0.049$. The chain was first run for 6000 MCMC steps with a 10-step COLA solver, after which it was switched to 20-steps and run for a further 9000 MCMC steps. The chain was converged by around 7000 MCMC steps, and we use samples from beyond 7000 steps for our results. The end product of the inference is a set of samples drawn from the posterior distribution for initial conditions consistent with the 2M++ galaxy distribution. Each sample consists of an initial density field at $z = 50$ on a 256^3 grid,¹ a final 256^3 redshift-space density field at $z = 0$, and the parameters of the bias model for each of the 16 galaxy catalogues.

2.2 Posterior resimulation

The second stage of our cluster mass estimation method requires resimulating many initial conditions sampled from the posterior distribution obtained using field-level inference, which can itself be a computational demanding task. However, it is not necessary to resimulate every sample from the Markov chain. One can resimulate a selection of samples from the chain and histogram the mass

estimates thus obtained, assuming the estimates are approximately independent. From these, we can compute the mean and variance of the estimated distribution.

In this work, we take 20 initial conditions from the chain run in Section 2.1, each separated by 300 MCMC steps (longer than the measured correlation length of any relevant parameter or field value). The initial conditions are generated with `geneticIC` (Stopyra et al. 2021a) from the 256^3 grid white-noise output of BORG, and over-sampled with `geneticIC`'s tricubic interpolation to generate 512^3 particles with a mass resolution of $2 \times 10^{11} M_{\odot} h^{-1}$, in order to reduce shot noise. They are then evolved from $z = 50$ to 0 with GADGET2 (Springel 2005) on a $677.7 h^{-1} \text{ Mpc}$ box to give a set of 20 simulations that sample the posterior distribution of the local density field.

To find haloes, we use the AHF halo finder (Knollmann & Knebe 2009), which identifies spherical-overdensity haloes in the resimulations, and except where otherwise stated we adopt the M_{200c} mass definition (*i.e.* the mass enclosed within a sphere whose mean density is 200 times the critical density of the Universe).

For each sample, we also perform simulations with inverted initial conditions to study the abundance of ‘antihaloes’ as a probe of voids (Pontzen et al. 2016). Antihaloes are a model of voids defined in N -body simulations by reversing the density-contrast of the initial conditions (swapping under- and overdensities), evolving the reversed initial conditions to redshift zero, and mapping the halo particles in the resulting ‘anti-universe’ simulation into the original simulation. Once mapped into the original simulation, they correspond to voids, with the benefit that their abundance is closely

¹Note that Jasche & Lavaux (2019) used $z = 69$ as their initial redshift.

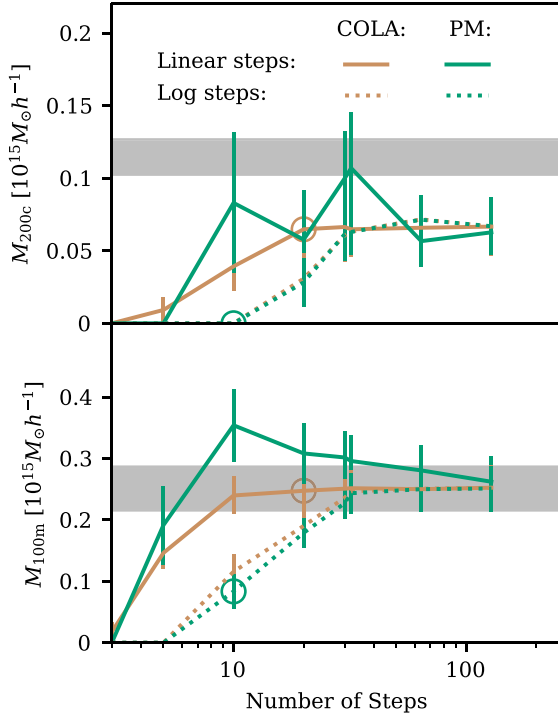


Figure 2. Convergence of mass in the same simulations as Fig. 1, but for a lower mass cluster with $M_{200c} \simeq 10^{14} M_{\odot} h^{-1}$. As for the higher mass examples, the COLA forward model is able to reproduce the M_{100m} mass; however, due to limitations in spatial resolution it is unable to reproduce M_{200c} , regardless of number of time-steps. This necessitates posterior resimulation of initial conditions with GADGET if M_{200c} masses are to be accurately recovered.

related to that of haloes. Posterior resimulation is particularly well-suited to studying antihaloes, since the initial conditions are available for inversion and resimulation.

2.3 Posterior predictive tests

To verify the accuracy of the inferred initial conditions, we perform posterior predictive tests to compare the posterior-predicted galaxy counts to those found in the 2M++ galaxy catalogue. For this task, we require the posterior distribution on the expected number of galaxies in the i th voxel, λ_i , given observed 2M++ galaxy counts. Obtaining this is complicated by the fact that (see Appendix A), the likelihood used in this work marginalizes over the amplitudes A_{α}^c in equation (2), and hence this posterior is not explicitly provided by the BORG MCMC chain. However, the required posterior can be constructed from the MCMC chain, since it provides samples from the posterior on $\bar{\lambda}_i^c \equiv \lambda_i^c / A_{\alpha(i)}^c$. We find (see Appendix B for details) that the expectation value of A_{α}^c for catalogue c is given by

$$E(A_{\alpha}^c | N_{1\dots I}^c) \simeq \frac{N_{\text{tot},\alpha}^c}{S} \sum_{s=1}^S \frac{1}{\bar{\lambda}_{\text{tot},\alpha,s}^c}, \quad (4)$$

where $i = \{1\dots I\}$ are assumed to be the unmasked voxels in the healpixel α , s indexes S samples from the posterior, $N_{\text{tot},\alpha}^c = \sum_{i=1}^I N_i^c$, and $\bar{\lambda}_{\text{tot},\alpha,s}^c = \sum_{i=1}^I \bar{\lambda}_{i,s}^c$. Furthermore, the posterior distribution $P(\lambda_i^c | N_{1\dots I}^c) = P(A_{\alpha}^c \bar{\lambda}_i^c | N_{1\dots I}^c)$ has expectation value

$$E(A_{\alpha}^c \bar{\lambda}_i^c | N_{1\dots I}^c) \simeq \frac{N_{\text{tot},\alpha}^c}{S} \sum_{s=1}^S \frac{\bar{\lambda}_{i,\alpha,s}^c}{\bar{\lambda}_{\text{tot},\alpha,s}^c}. \quad (5)$$

Given that the set up using multiple amplitudes and catalogues makes the correlation structure of the inferred field and hence its variance cumbersome to calculate analytically, we proceed using a bootstrap estimate of the uncertainty on the 2M++ data. In particular, for each of the 16 galaxy catalogues, we bootstrap the 2M++ galaxy counts for the voxels in each spherical shell to obtain an estimate of the uncertainty on the total number of galaxies in that shell.

3 RESULTS

This section is structured as follows: in Section 3.1, by considering the convergence of mass estimates as a function of physical scale, we show that the COLA20 gravity solver is adequate for obtaining reliable mass estimates for the highest mass clusters. In Section 3.2, we present the results for mass functions obtained using COLA20, and compare it to the previous state-of-the-art PM10 inference. In Section 3.3, we discuss how individual clusters with masses between 10^{14} and $10^{15} M_{\odot} h^{-1}$ can be identified within the posterior resimulations, comparing the results with a collection of mass estimates from the literature. Finally, we present the results of posterior predictive tests for the galaxy counts in these clusters, noting some possible indications of remaining systematic uncertainties (Section 3.4).

3.1 Choice of gravity solver

We begin by establishing the accuracy needed for the gravity solver within the field-level inference in order to obtain reliable cluster masses. This is accomplished by testing which gravity solvers can predict converged masses relative to a GADGET2 resimulation, starting from the same initial conditions. We used initial conditions from the PM10 MCMC chain Jasche & Lavaux (2019) at $z = 69$ and evolved them to $z = 0$ using a variety of solvers with different, fixed numbers of time-steps. We identified massive clusters in the simulations by searching for the largest halo within a $20h^{-1}$ Mpc radius of the known position of a cluster in the Abell catalogue (Abell, Corwin & Olowin 1989). We examined two different definitions of mass (M_{200c} , the mass contained within a sphere of density 200 times the critical density, and M_{100m} , the mass contained within a sphere of density 100 times the mean density of the Universe). This allowed us to examine how the gravity solvers perform at different scales: M_{200c} virial radii are typically half that of M_{100m} virial radii. Convergence to the correct mass for a given set of initial conditions indicates that the solver is consistent with N -body simulations at a particular scale.

We show the results of the time-step tests for several different clusters in the Jasche & Lavaux (2019) MCMC chain in Fig. 1 for high mass, $\sim 10^{15} M_{\odot} h^{-1}$ clusters.² In each case, we consider a range of step numbers between 3 and 128, and we also consider linear or logarithmic spacing of these steps in scale factor between $z = 69$ and 0. Grey bands show the standard deviation of the mean mass for these haloes over the samples taken from the posterior, while the lines show the mean mass over all contributing MCMC samples (error bar is the standard deviation of the mean).

At this mass scale, all the gravity solvers we considered converged to the same mass as GADGET simulations when the number of time-steps is increased, representing the increased accuracy (but also increased computational cost) that comes with using more time-steps.

²Note that because these tests were done with initial conditions from the Jasche & Lavaux (2019) MCMC chain which used a PM10 gravity solver, the masses are not expected to be consistent with the masses found for these clusters in Fig. 4.

For the COLA model, 10–20 time-steps were found to be sufficient to reproduce the mass of the largest clusters ($\sim 10^{15} M_{\odot} h^{-1}$) at a level consistent with GADGET (see Fig. 1).

We repeated our test for lower-mass clusters ($\sim 10^{14} M_{\odot} h^{-1}$), with an example being shown in Fig. 2. In these cases, increasing the time-stepping accuracy does not result in convergence to the GADGET result for M_{200c} . This implies that the primary limitation on the accuracy of the gravity solver is spatial rather than temporal resolution on these scales. The grid scale is $2.65 h^{-1}$ Mpc, considerably larger than the virial radius $R_{200c} \simeq 0.75 h^{-1}$ Mpc for this definition. However, convergence is still achieved for the M_{100m} masses, and at this mass scale $R_{100m} \simeq 1.4 h^{-1}$ Mpc. This implies that the overall mass in the cluster environment can be correctly reproduced with an approximate gravity solver, provided one does not extrapolate too far below the grid scale.³

Choosing a specific gravity solver, time-step number, and time-step spacing implies a trade-off between accuracy and runtime. Ideally, one would choose a time-step that achieves convergence for spatially-resolved quantities. Based on the results discussed above, we chose COLA with 20 linearly-spaced steps as the best trade-off for the spatial resolution of the inference here. While we did not consider other integrators such as FASTPM in this work, recent work by List & Hahn (2023) suggests that the incorporation of Lagrangian perturbation theory information into an integrator allows for optimal use of time-steps. We therefore expect that FASTPM should perform similarly well to COLA for this purpose.

We emphasize that, while spatial resolution prevents M_{200c} convergence for low-mass clusters ($\sim 10^{14} h^{-1}$ Mpc) within the BORG gravity solver, resimulations with GADGET starting from posterior samples will overcome this limitation. Based on the understanding that M_{200c} is dictated by the larger-scale flows that are resolved, it is therefore legitimate to test whether the resimulated halo mass functions agree with expectations.

3.2 Mass functions

In Fig. 3, we show the halo and antihalo M_{200c} mass functions inferred when using posterior resimulation applied to samples from BORG with the COLA20 gravity solver, and compare this with the same results using the previous PM10-based inference.

Fig. 3 shows that the COLA20 mass functions are in agreement with those obtained from regions of similar underdensity in unconstrained simulations (shown by the blue-shaded region), and are therefore compatible with Λ CDM expectations. Conversely, when using the PM10 gravity solver from Jasche & Lavaux (2019), the resimulated mass functions overpredict the halo and antihalo abundances. To obtain ‘similar underdensity’ regions in unconstrained simulations, we first computed the mean density contrast of the central $135 h^{-1}$ Mpc region over all 20 MCMC samples from the COLA20 chain, which gave $\delta = -0.043 \pm 0.001$ (standard error of the mean). We then select $135 h^{-1}$ Mpc spheres centred randomly within a set of unconstrained simulations with cosmological parameters matching that of the BORG inference, and retain those spheres whose density contrast lies within one standard deviation of this mean ($-0.044 < \delta < -0.042$).

³Note that recovering the M_{200c} mass of haloes is generally a more stringent test than, for example, reproducing the Friend-of-Friends halo-mass function with linking length $l = 0.2 h^{-1}$ Mpc, which can be done with approximate solvers such as COLA or FASTPM (Feng et al. 2016) at the per cent level (Izard, Crocce & Fosalba 2016).

These results show that insufficient integration accuracy leads the sampler to push linear over- and underdensities to exaggerated values in compensation; when resimulated, this leads to unphysically high mass clusters and antihaloes. In more detail, the PM10 solver presented by Jasche & Lavaux (2019) has insufficient time-steps at low redshift to properly resolve the final-stage collapse of even high-mass haloes. As a result, it underestimates the true density at the core of massive haloes (Fig. 1); this causes BORG to overestimate the initial conditions in order to infer the correct final density field. This results in inflated cluster masses when the initial conditions are resimulated with more accurate N -body solvers. This indirectly affects the masses of the antihaloes as well, since consistency with the larger-scale underdensity of the local super-volume requires the extra mass in these clusters to be taken from surrounding regions, resulting in an excess of antihaloes.

These results explain the excess of clusters found by SIBELIUS-DARK (McAlpine et al. 2022) and Hutt et al. (2022), as well as the excess of antihaloes relative to Λ CDM noted in Desmond et al. (2022); all of these works used the Jasche & Lavaux (2019) inference based on 2M++. Our COLA20-based result demonstrates that the excess of antihaloes and haloes is not a real effect, and disappears when a more accurate gravity solver is used for the inference. The requirements for reconstructing *individual* cluster masses, however, may be even more stringent than the requirements for the mass function, and we now turn to this issue.

3.3 Individual cluster masses

In the resimulations based on COLA20, we again identify clusters by searching for the largest halo within a $20 h^{-1}$ Mpc radius, as we previously applied to the older PM10-based resimulations (Section 3.1). This leads to an unambiguous identification of the relevant halo for clusters of masses approaching $10^{15} M_{\odot} h^{-1}$, which are well-constrained by the field-level inference. There are nine such cases, and we identify the clusters as Perseus–Pisces (A426), Hercules B (A2147), Coma (A1656), Norma (A3627), Shapley (A3571), A548, Hercules A (A2199), Hercules C (A2063), and Leo (A1367). The mean mass of each cluster using the posterior distribution is obtained by averaging the halo masses for all its counterparts across all 20 resimulations.

To compare these M_{200c} mass estimates to known data, we make use of previously collected mass estimates for nine nearby massive clusters. The estimates come from dynamical, X-ray, SZ and weak lensing techniques; full details of the estimates are discussed in Stopyra et al. (2021b). These mass estimates are shown in Fig. 4 and compared to our resimulation estimates using the new COLA20-based field-level inference. In most cases, these new results are consistent with existing mass estimates. Perseus–Pisces (A426) is a notable exception, which we will return to in Section 3.4.

For several of these clusters, M_{200c} is in the order of 68 per cent. As previously discussed in relation to Fig. 2, the gravity solver used within the inference is prevented by its spatial resolution from directly predicting such masses; the resimulations, however, remove this limitation.

3.4 Posterior predictive tests

So far, we have shown that using COLA20 as the gravity integrator within the BORG pipeline, then resimulating to obtain a final cluster (or antihalo) catalogue, gives rise to M_{200c} cluster and antihalo mass functions in good agreement with Λ CDM expectations. We have further shown that the mass estimates for individual clusters obtained

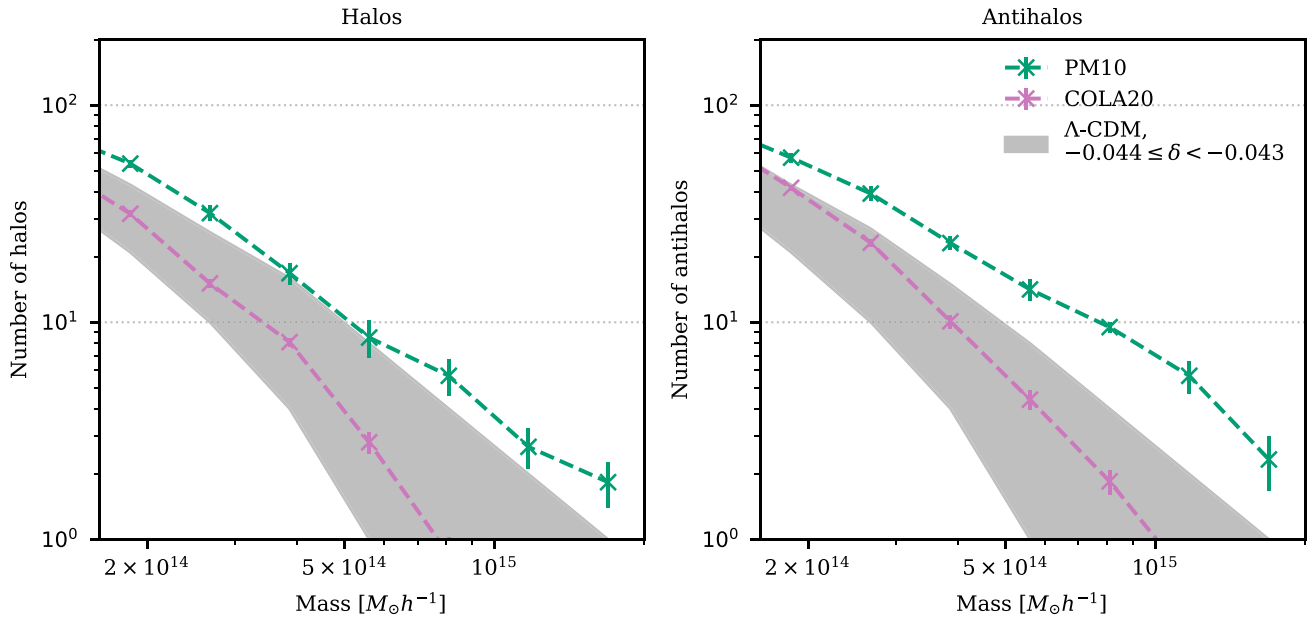


Figure 3. Halo mass function (left) and antihalo mass function (right) for the central $135 h^{-1}$ Mpc region (covering the local super-volume) obtained using posterior resimulation of the new BORG COLA20 field-level inference (pink lines, this work) versus the original PM10 inference (Jasche & Lavaux 2019) (green lines). The shaded regions show the 95 per cent Poisson interval expected using mass functions estimated from GADGET simulations with the same cosmological parameters as the COLA20-based inference, conditioned on regions with a density contrast matches that of the local super-volume ($\delta = -0.043 \pm 0.001$).

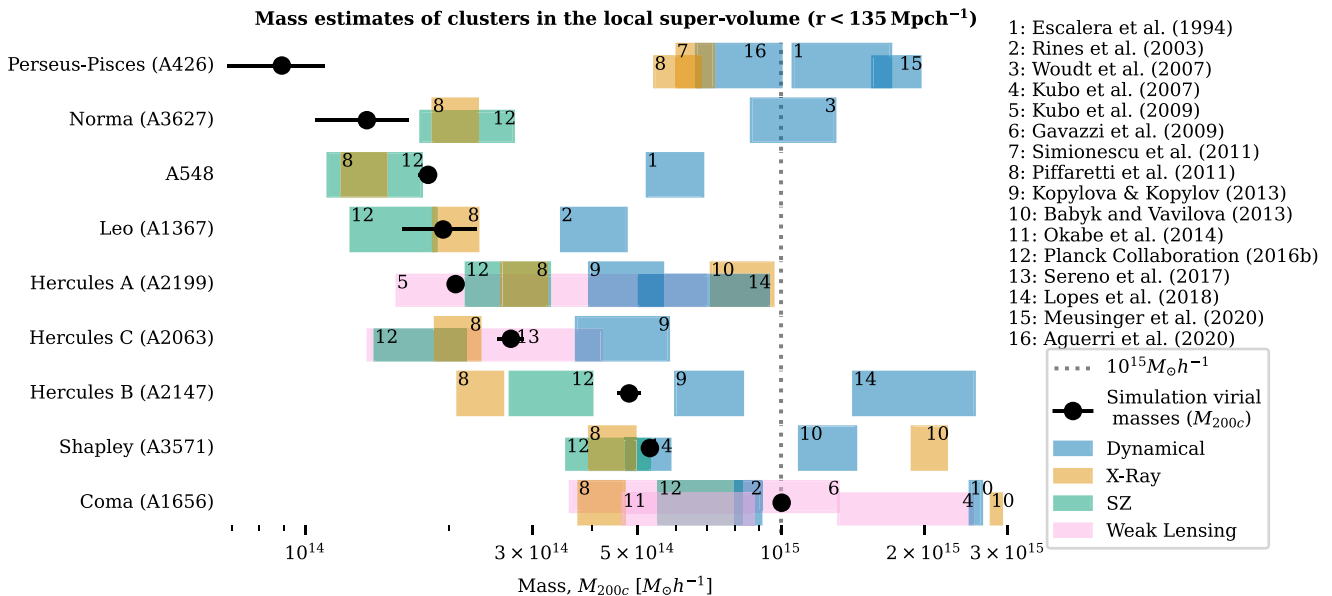


Figure 4. Cluster mass estimates (M_{200c}) obtained with the resimulated COLA20 BORG inference (black dots), compared with other M_{200c} mass estimates for the same clusters as compiled by Stopyra et al. (2021b) (non M_{200c} estimates are converted M_{200c} via a concentration-mass relationship). The uncertainties on the BORG estimates are given by the standard deviation of the distribution of halo masses associated with a given cluster, obtained from 20 resimulated MCMC samples. In most cases, the masses are consistent with other estimates. Perseus–Pisces (A426) is a notable exception, which we discuss in Section 3.4.

in this way are, in most cases, in agreement with independent observational estimates.

We next computed the posterior predictive tests outlined in Section 2.3 for the nine individual massive clusters discussed in the previous section. Since the majority of the clusters pass the posterior

predictive test, we show in Fig. 5 just two illustrative examples: Coma and Perseus–Pisces, which contain similar numbers of galaxies. Posterior predictive tests for the other seven clusters are shown in Appendix C. The dark- and light-shaded regions of Fig. 5 indicate 68 and 95 per cent credible intervals for the 2M++ galaxy counts

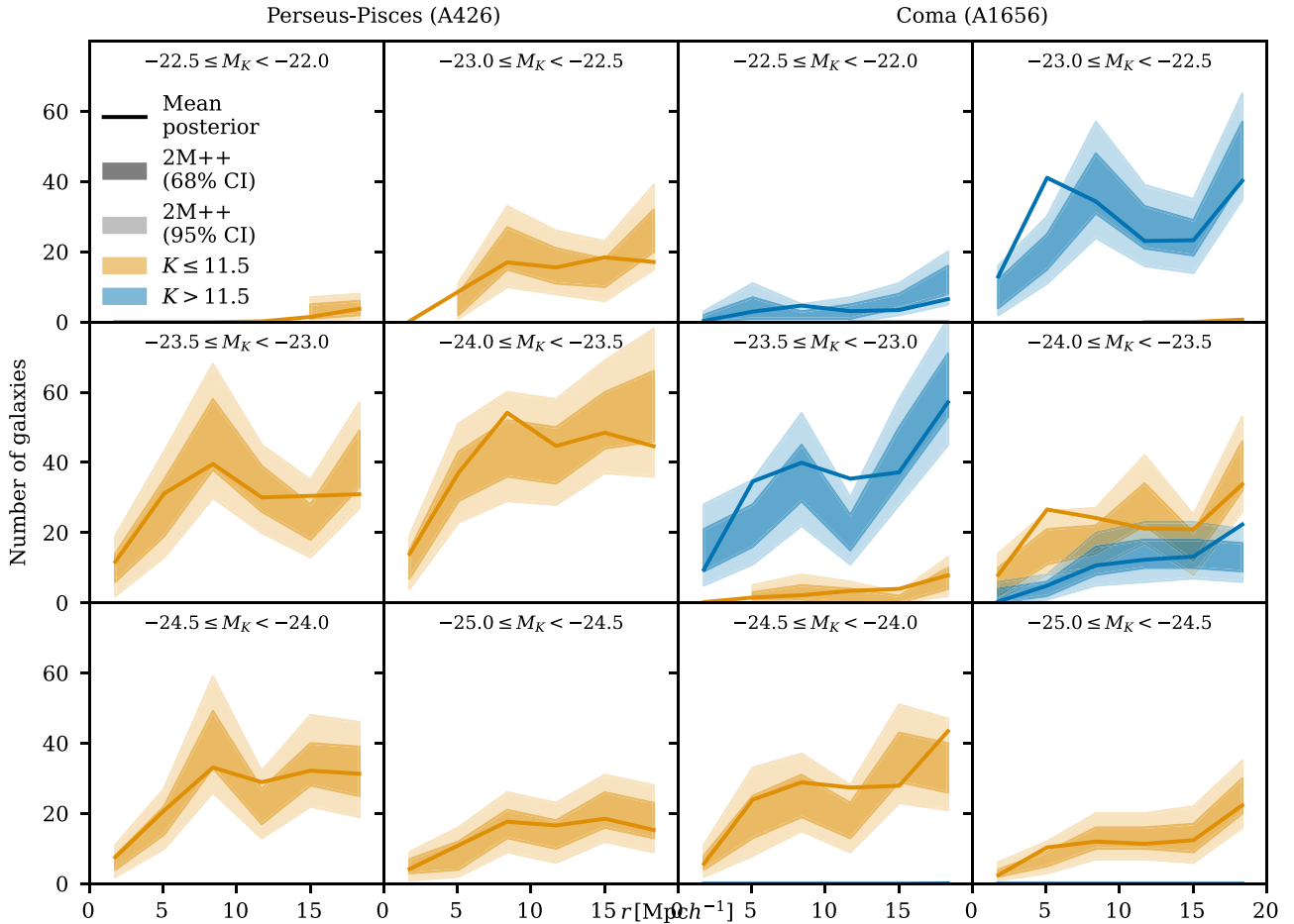


Figure 5. Posterior predictive tests for the galaxy count in radial 68 per cent-wide shells around two clusters in the local super-volume: Perseus–Pisces (left-hand panels) and Coma (right-hand panels). Solid lines show the predicted mean counts from the posterior distribution for each absolute magnitude bin, while shaded regions show the 68 and 95 per cent credible intervals computed by bootstrapping the sum of all voxels in each shell for the 2M++ galaxies. The $K \leq 11.5$ catalogue is shown in orange, while the $K > 11.5$ catalogue is shown in blue. Note that Perseus–Pisces entirely lacks $K > 11.5$ catalogue data due to being in the 2MRS portion of the sky. Despite its apparently underestimated mass in Fig. 4, the posterior predictive tests pass in all magnitude bins.

in radial shells around the centres of each cluster, estimated using bootstrap. The solid lines show the mean of the posteriority predicted galaxy counts in the same shells computed using equation (5). Orange and blue colours, respectively, indicate the $K \leq 11.5$ and $K > 11.5$ catalogues.

In the case of Perseus–Pisces, while the mass appears substantially underestimated relative to constraints in the literature (as shown in Fig. 4), the posterior-predicted number of galaxies is consistent with the available data. This raises the question of how the mass can be so low compared to Coma, which has a similar number of observed galaxies but an order of magnitude higher inferred mass. The connection from the inferred density field to predicted number counts is dictated by a global bias model per catalogue, but it is additionally locally modulated in each healpix pixel by an amplitude A_α^c . We therefore investigated the expectation value of A_α^c for the healpix in the $10 h^{-1}$ Mpc surrounding Coma and Perseus–Pisces to test whether these amplitudes can account for the differing results.

Fig. 6 shows, in absolute magnitude bins (left to right) for Perseus–Pisces (top) and Coma (bottom), the distribution of the expectation values for healpixel amplitudes across the entire local

super-volume. These are computed using equation (4), and $K \leq 11.5$ and $K > 11.5$ results are, respectively, shown as orange and blue histograms. Vertical arrows show the amplitudes for healpixels within $10 h^{-1}$ Mpc of the centre of each cluster. First, we note that Perseus–Pisces has no galaxies in the $K > 11.5$ catalogues (and therefore these histograms are not shown). Second, in the bright ($K \leq 11.5$) galaxies, the amplitudes inferred around Perseus–Pisces are preferentially in the high-end tail. This is indicative of a potential systematic error, allowing an unphysically small overdensity to generate a large number of galaxies. However, obtaining further insight into this result requires a detailed study of the interaction between the local amplitudes and the global bias models, or even a revision of the construction of the original catalogue, which is beyond the scope of the present work. We also note that Perseus–Pisces appears to lack SZ or weak lensing mass estimates in the literature, and that compiled dynamical mass estimates for other clusters are nearly all overestimates relative to other methods. In summary, there are strong indications that our posterior resimulation mass inference for Perseus–Pisces is biased, but we also cannot rule out the possibility that mass estimates using other methods for Perseus–Pisces may require revision.

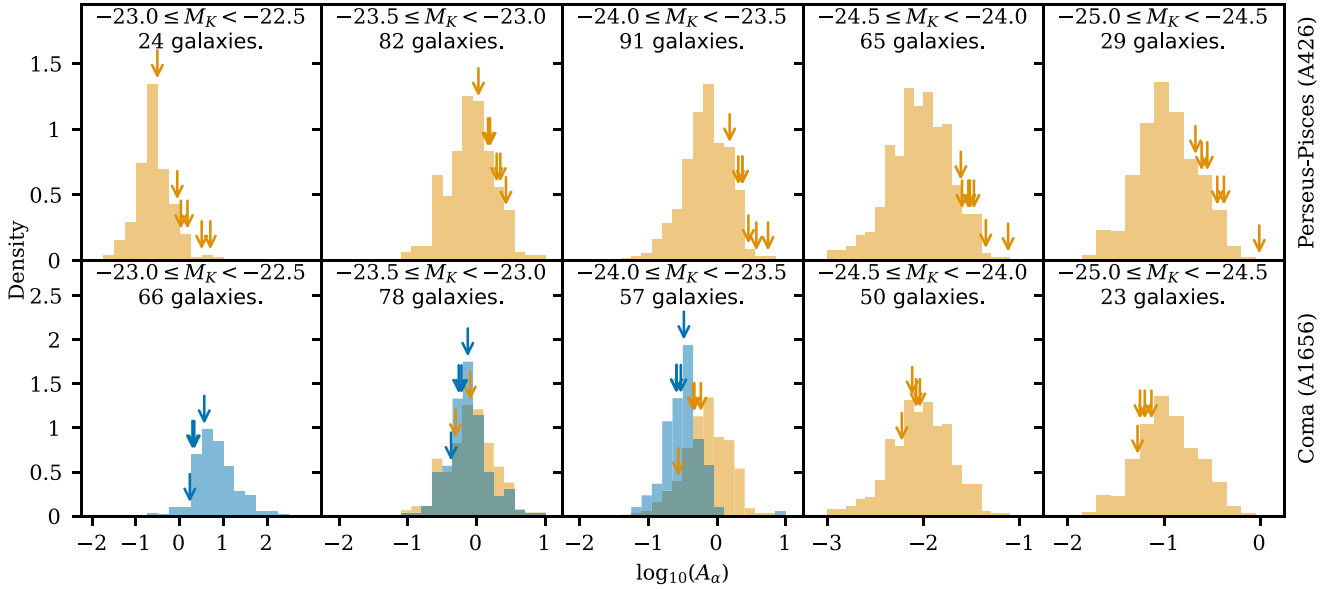


Figure 6. Distribution across the full sky of expected amplitudes, A_α , given by equation (4) for each of five magnitude bins (histograms), with arrows indicating the amplitudes for the healpixels within $10 h^{-1}$ Mpc of the centre of the two clusters Perseus–Pisces (top) and Coma (bottom). Orange again indicates the $K \leq 11.5$ catalogue, while blue indicates the $11.5 < K \leq 12.5$ catalogue. The number of 2M++ galaxies within $10 h^{-1}$ Mpc of the cluster centre, which lie in each absolute magnitude bin are specified. The healpixel amplitudes for Perseus–Pisces consistently lie in the high-amplitude tail of the distribution. Note that the distance to Perseus–Pisces ($54.6 h^{-1}$ Mpc) places it close to the boundary between two different $60 h^{-1}$ Mpc-wide healpix shells. It therefore receives contributions from many more healpix regions than Coma.

4 DISCUSSION

In a previous work, we showed that the number of clusters and antihaloes in the local super-volume with masses above $10^{15} M_\odot h^{-1}$ is a powerful test of consistency with Λ CDM (Stopyra et al. 2021b). Related cosmological tests include assessments of the rarity of structures such as the Sloan Great Wall and the Shapley supercluster (Nichol et al. 2006; Sheth & Diaferio 2011). While the 2M++ catalogue does not have sufficient signal-to-noise to directly assess the rarity of the Sloan Great Wall, we find that the number of massive clusters in Fig. 4 and the mass function in Fig. 3 appear compatible with Λ CDM, with masses broadly consisting of X-ray and SZ estimates. Our results show that the number of massive clusters and antihaloes in the local super-volume is compatible with theoretical expectations from a Λ CDM universe.

The question of whether the local super-volume lies at the centre of a large Local Void has attracted much attention in the literature (Frith et al. 2003; Xie, Gao & Guo 2014; Shanks et al. 2019b). More recently, it has been suggested that a large-scale underdensity might be responsible for the Hubble tension (Shanks, Hogarth & Metcalfe 2019a). In particular, the 2MASS survey has been claimed to provide evidence for such a large-scale underdensity (Frith et al. 2003; Shanks et al. 2019b). Using field-level inference with BORG, which includes 2MASS data as part of the 2M++ catalogue, we are able to directly probe the dark-matter density field and thus infer the size of the large-scale underdensity. We find that the local super-volume has a density contrast of $\delta = -0.043 \pm 0.001$ out to a radius of $135 h^{-1}$ Mpc. While this means that the local super-volume is underdense, it is not sufficiently underdense to have a significant impact on measurements of the Hubble rate (Wu & Huterer 2017), so does not help by itself to alleviate the Hubble tension.

5 CONCLUSIONS

In this work, we assessed different gravity solvers that can be used for field-level inference in the context of the BORG algorithm followed by posterior resimulation. We showed that replacing the 10-step PM solver used in the Jasche & Lavaux (2019) with a 20-step COLA solver corrects the overabundance of massive haloes and voids previously noted by McAlpine et al. (2022), Hutt et al. (2022), Desmond et al. (2022).

We used the 20-step COLA solver to construct a new BORG inference from the 2M++ catalogue, with updated cosmological parameters and likelihood. The new inference, combined with posterior resimulation, leads to halo and void mass functions, which are consistent with Λ CDM expectations, and the masses of individual massive local clusters are broadly consistent with other estimates in the literature. An exception to this, however, was the Perseus–Pisces cluster, which appeared to have a significantly lower mass than expected from its galaxy counts and from other estimates of its mass in the literature. We discussed possible reasons for this in Section 3.4.

Our results show that the use of field-level inference for extracting information on non-linear scales requires careful control of the accuracy of the forward modelling used. We have shown how to validate the forward modelling for the accurate estimation of cluster masses using posterior resimulation. In the process, we have created a new inference of the initial conditions consistent with the local Universe and provided constraints on the local underdensity, and the number of massive nearby clusters. In future work, we will apply this inference to constructing an anti-halo void catalogue of the local super-volume.

ACKNOWLEDGEMENTS

We thank Daniel Mortlock, Harry Desmond, Stuart McAlpine, and Metin Ata for useful discussions regarding this work. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programmes (grant agreement no. 101018897 CosmicExplorer and 818085 GMGalaxies). This work has been enabled by support from the research project grant ‘Understanding the Dynamic Universe’ funded by the Knut and Alice Wallenberg Foundation under Dnr KAW 2018.0067. SS and HVP are additionally supported by the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine. HVP and JJ acknowledge the hospitality of the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. The participation of HVP and JJ at the Aspen Center for Physics was supported by the Simons Foundation. JJ acknowledges support by the Swedish Research Council (VR) under the project 2020-05143 – ‘Deciphering the Dynamics of Cosmic Structure’ and by the Simons Collaboration on ‘Learning the Universe’. This work was partially enabled by the UCL Cosmoparticle Initiative. The computations/data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Linköping University, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. This research also utilized the Sunrise HPC facility supported by the Technical Division at the Department of Physics, Stockholm University. This work was carried out within the Aquila Consortium.⁴

AUTHOR CONTRIBUTIONS

The first four authors contributed roughly equally to this work, with differing areas of contribution. GL developed and implemented the τ COLA gravity solver used by BORG, and set up the COLA20 chain used in this analysis. We outline the different contributions below using key words based on the CRediT (Contribution Roles Taxonomy) system.

SS: data curation; investigation; formal analysis; software; visualization; writing – original draft preparation.

HVP: conceptualization; methodology; validation, and interpretation; writing (original draft; review and editing); funding acquisition.

AP: conceptualization; methodology; validation and interpretation; writing (original draft; review, and editing).

JJ: methodology; data curation; resources; software; validation, and interpretation; writing – review

GL: data curation; resources; software; writing – review.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Abazajian K. N. et al., 2009, *Astrophys. J. Suppl.*, 182, 543
 Abell G. O., Corwin Harold G. J., Olowin R. P., 1989, *ApJS*, 70, 1
 Aguerri J. A. L., Girardi M., Agulli I., Negri A., Dalla Vecchia C., Domínguez Palmero L., 2020, *MNRAS*, 494, 1681
 Ata M., Lee K.-G., Vecchia C. D., Kitaura F.-S., Cucciati O., Lemaux B. C., Kashino D., Müller T., 2022, *Nature Astron.*, 6, 857
 Babić I., Schmidt F., Tucci B., 2022, *J. Cosmol. Astropart. Phys.*, 08, 007
 Babyk I. V., Vavilova I., 2013, *Odessa Astron. Publ.*, 26, 175
 Bertschinger E., 1985, *ApJS*, 58, 39

- Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, *MNRAS*, 456, 2361
 Codis S., Pichon C., Pogosyan D., 2015, *MNRAS*, 452, 3369
 Costanzi M. et al., 2019, *MNRAS*, 488, 4779
 Desmond H., Hutt M. L., Devriendt J., Slyz A., 2022, *MNRAS*, 511, L45
 Doumler T., Hoffman Y., Courtois H., Gottlöber S., 2013, *MNRAS*, 430, 888
 Duane S., Kennedy A. D., Pendleton B. J., Roweth D., 1987, *Phys. Lett. B*, 195, 216
 Eastwood J., Hockney R., 1974, *J. Comput. Phys.*, 16, 342
 Escalera E., Biviano A., Girardi M., Giuricin G., Mardirossian F., Mazure A., Mezzetti M., 1994, *ApJ*, 423, 539
 Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, *MNRAS*, 463, 2273
 Frith W. J., Buswell G., Fong R., Metcalfe N., Shanks T., 2003, *MNRAS*, 345, 1049
 Gavazzi R., Adami C., Durret F., Cuillandre J. C., Ilbert O., Mazure A., Pelló R., Ulmer M. P., 2009, *A&A*, 498, L33
 Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
 Gottlöber S., Hoffman Y., Yepes G., 2010, in Wagner S., Bode A., Müller M. M., Steinmetz M., eds, *High Performance Computing in Science and Engineering, Garching/Munich 2009*. Springer, Heidelberg, p. 309
 Heß S., Kitaura F.-S., Gottlöber S., 2013, *MNRAS*, 435, 2065
 Hoffman Y., Ribak E., 1991, *ApJ*, 380, L5
 Horowitz B., Lee K.-G., White M., Krolewski A., Ata M., 2019, *ApJ*, 887, 61
 Huchra J. P. et al., 2012, *ApJS*, 199, 26
 Hutt M. L., Desmond H., Devriendt J., Slyz A., 2022, *MNRAS*, 516, 3592
 Ilić S., Sakr Z., Blanchard A., 2019, *A&A*, 631, A96
 Izard A., Crocce M., Fosalba P., 2016, *MNRAS*, 459, 2327
 Jasche J., Lavaux G., 2019, *A&A*, 625, A64
 Jasche J., Wandelt B. D., 2013, *MNRAS*, 432, 894
 Jones D. H., Peterson B. A., Colless M., Saunders W., 2006, *MNRAS*, 369, 25
 Klypin A. A., Shandarin S. F., 1983, *MNRAS*, 204, 891
 Klypin A., Holtzman J., 1997, preprint (astro:–ph/9712217)
 Knollmann S. R., Knebe A., 2009, *ApJS*, 182, 608
 Kopylova F. G., Kopylov A. I., 2013, *Astron. Lett.*, 39, 1
 Kostić A., Nguyen N.-M., Schmidt F., Reinecke M., 2023, *J. Cosmol. Astropart. Phys.*, 07, 063
 Kubo J. M. et al., 2009, *ApJ*, 702, L110
 Kubo J. M., Stebbins A., Annis J., Dell’Antonio I. P., Lin H., Khiabani H., Frieman J. A., 2007, *ApJ*, 671, 1466
 Lavaux G., 2010, *MNRAS*, 406, 1007
 Lavaux G., Hudson M. J., 2011, *MNRAS*, 416, 2840
 Libeskind N. I. et al., 2020, *MNRAS*, 498, 2968
 List F., Hahn O., 2023, preprint (arXiv:2301.09655)
 Lopes P. A. A., Trevisan M., Laganá T. F., Durret F., Ribeiro A. L. B., Rembold S. B., 2018, *MNRAS*, 478, 5473
 LoVerde M., Smith K. M., 2011, *J. Cosmology Astropart. Phys.*, 08, 003
 Lucie-Smith L., Adhikari S., Wechsler R. H., 2022, *MNRAS*, 515, 2164
 Lucie-Smith L., Peiris H. V., Pontzen A., 2019, *MNRAS*, 490, 331
 Mak D. S., Pierpaoli E., Schmidt F., Macellari N., 2012, *Phys. Rev. D*, 85, 123513
 McAlpine S. et al., 2022, *MNRAS*, 512, 5823
 Meusinger H., Rudolf C., Stecklum B., Hoeft M., Mauersberger R., Apai D., 2020, *A&A*, 640, A30
 Neal R. M., 1993, in *Probabilistic inference using Markov chain Monte Carlo methods*. Department of Computer Science, University of Toronto, Toronto, ON, Canada
 Neyrinck M. C., Aragón-Calvo M. A., Jeong D., Wang X., 2014, *MNRAS*, 441, 646
 Nguyen N.-M., Schmidt F., Lavaux G., Jasche J., 2021, *J. Cosmology Astropart. Phys.*, 03, 058
 Nichol R. C. et al., 2006, *MNRAS*, 368, 1507
 Okabe N., Futamase T., Kajisawa M., Kuroshima R., 2014, *ApJ*, 784, 90
 Peebles P. J. E., 1989, *ApJ*, 344, L53
 Piffaretti R., Arnaud M., Pratt G. W., Pointecouteau E., Melin J. B., 2011, *A&A*, 534, A109
 Planck Collaboration et al., 2016, *A&A*, 594, A27
 Planck Collaboration et al., 2020, *A&A*, 641, A6

⁴<https://aquila-consortium.org>

- Pontzen A., Slosar A., Roth N., Peiris H. V., 2016, *Phys. Rev. D*, 93, 103519
- Porqueres N., Heavens A., Mortlock D., Lavaux G., 2022, *MNRAS*, 509, 3194
- Porqueres N., Heavens A., Mortlock D., Lavaux G., Makinen T. L., 2023, preprint (arXiv:2304.04785)
- Porqueres N., Jasche J., Lavaux G., Enßlin T., 2019b, *A&A*, 630, A151
- Porqueres N., Ramanah D. K., Jasche J., Lavaux G., 2019a, *A&A*, 624, A115
- Pratt G., Arnaud M., Biviano A., Eckert D., Ettori S., Nagai D., Okabe N., Reiprich T., 2019, *Space Sci. Rev.*, 215, 1
- Rines K., Geller M. J., Kurtz M. J., Diaferio A., 2003, *AJ*, 126, 2152
- Sartoris B., Borgani S., Fedeli C., Matarrese S., Moscardini L., Rosati P., Weller J., 2010, *MNRAS*, 407, 2339
- Sawala T., McAlpine S., Jasche J., Lavaux G., Jenkins A., Johansson P. H., Frenk C. S., 2022, *MNRAS*, 509, 1432
- Sereno M., Covone G., Izzo L., Ettori S., Coupon J., Lieu M., 2017, *MNRAS*, 472, 1946
- Shanks T., Hogarth L., Metcalfe N., 2019a, *MNRAS*, 484, L64
- Shanks T., Hogarth L., Metcalfe N., Whitbourn J., 2019b, *MNRAS*, 490, 4715
- Sheth R. K., Diaferio A., 2011, *MNRAS*, 417, 2938
- Simionescu A. et al., 2011, *Science*, 331, 1576
- Sorce J. G. et al., 2016, *MNRAS*, 455, 2078
- Sorce J. G., Courtois H. M., Gottlöber S., Hoffman Y., Tully R. B., 2014, *MNRAS*, 437, 3586
- Springel V., 2005, *MNRAS*, 364, 1105
- Stopyra S., Peiris H. V., Pontzen A., Jasche J., Natarajan P., 2021b, *MNRAS*, 507, 5425
- Stopyra S., Pontzen A., Peiris H., Roth N., Rey M. P., 2021a, *ApJS*, 252, 28
- Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, *J. Cosmol. Astropart. Phys.*, 06, 036
- Woudt P. A., Kraan-Korteweg R. C., Lucey J., Fairall A. P., Moore S. A. W., 2008, *MNRAS*, 383, 445
- Wu H.-Y., Huterer D., 2017, *MNRAS*, 471, 4946
- Xie L., Gao L., Guo Q., 2014, *MNRAS*, 441, 933

APPENDIX A: LIKELIHOOD

The new BORG inference we used in this paper uses an updated likelihood compared to the Jasche & Lavaux (2019) inference, first outlined by Porqueres et al. (2019a). The aim of this approach is to make the likelihood robust to unknown systematics in the underlying data.

To achieve this, the sky is first subdivided into 192 healpix pixels (Gorski et al. 2005), with $n_{\text{side}} = 4$. The extension of these healpixels into three dimensions is then split into 10 radial bins of depth $60 h^{-1}$ Mpc each, giving 1920 regions on the sky, which we denote *patches*. Note that this is a separate pixelization to the 256^3 cubic voxels into which we divide the simulation. Each voxel, which we label with Roman letters, i , lies within a given healpix patch (labelled by Greek letters, α), and there is a unique mapping $\alpha(i)$ from voxels to the healpix patch which contains them.

All the voxels that lie in a given patch have a shared amplitude, A_α^c , for catalogue c , which is regarded as a nuisance parameter and marginalized over in the likelihood. Once the gravity solver is specified, equation (2) gives the expected number of galaxies for catalogue c in voxel i , λ_i^c , as a function of the final density field, $\delta = G(\delta^{\text{IC}})$. We define $\bar{\lambda}_i^c = \lambda_i^c / A_{\alpha(i)}^c$, where $A_{\alpha(i)}^c$ is the systematics amplitude parameter corresponding to the patch containing voxel i . We can then write a Poisson likelihood for the number of galaxies in catalogue c , N_i^c , that lie in voxels $i \in \{1, \dots, I\}$ in patch α as

$$P(N_{1\dots I}^c | \bar{\lambda}_{1\dots I}^c, A_\alpha^c) = \prod_{i=1}^I \frac{(A_\alpha^c \bar{\lambda}_i^c)^{N_i^c}}{N_i^c!} e^{-A_\alpha^c \bar{\lambda}_i^c}. \quad (\text{A1})$$

Porqueres et al. (2019a) chose to marginalize over the amplitudes A_α^c to create a new ‘robust’ likelihood that reduces sensitivity to

spatially-varying systematics in the data,

$$P(N_{1\dots I}^c | \bar{\lambda}_{1\dots I}^c) = \int dA_\alpha^c P(A_\alpha^c) P(N_{1\dots I}^c | \bar{\lambda}_{1\dots I}^c, A_\alpha^c). \quad (\text{A2})$$

They chose a Jeffreys prior $P(A_\alpha^c) = \kappa_\alpha / A_\alpha^c$, with normalization constant κ_α . This yields the marginalized likelihood

$$P(N_{1\dots I}^c | \bar{\lambda}_{1\dots I}^c) \propto \prod_{i=1}^I \left(\frac{\bar{\lambda}_i^c}{\sum_{j=1}^I \bar{\lambda}_j^c} \right)^{N_i^c}. \quad (\text{A3})$$

Note that we assume all catalogues c are independent except via their dependence on the final density δ_i used to compute $\bar{\lambda}_i^c$ in equation (2). The likelihood for the number of galaxies in all catalogues and voxels is therefore given by the product of equation (A3) over all catalogues, c , and all patches, α .

APPENDIX B: POSTERIOR PREDICTIVE TESTS WITH THE ROBUST LIKELIHOOD

The fact that the amplitudes in equation (2) are marginalized over in the robust likelihood presented in Appendix A makes it more difficult to interrogate the inferred model using posterior predictive testing. We wish to compare the posterior-predicted number of galaxies to that actually found in the 2M++ catalogue in specific regions of interest; this requires a posterior for the amplitudes, A_α^c . We now show how to reconstruct the information we need for constructing posterior predictive tests using the information we do have access to.

BORG samples the posterior for initial density and bias parameters, which are combined to give $\bar{\lambda}_i^c = \lambda_i^c / A_{\alpha(i)}^c$. If we have a healpix region containing voxels $i = 1 \dots I$, with amplitude A_α^c in catalogue c , we can compute the joint posterior distribution for $\bar{\lambda}_i^c$ and A_α^c conditioned on $N_{1\dots I}^c$:

$$\begin{aligned} P(\bar{\lambda}_{1\dots I}^c, A_\alpha^c | N_{1\dots I}^c) &= P(\bar{\lambda}_{1\dots I}^c | N_{1\dots I}^c) P(A_\alpha^c | \bar{\lambda}_{1\dots I}^c, N_{1\dots I}^c) \\ &= P(\bar{\lambda}_{1\dots I}^c | N_{1\dots I}^c) \frac{P(A_\alpha^c | \bar{\lambda}_{1\dots I}^c) P(N_{1\dots I}^c | A_\alpha^c, \bar{\lambda}_{1\dots I}^c)}{P(N_{1\dots I}^c | \bar{\lambda}_{1\dots I}^c)} \\ &= P(\bar{\lambda}_{1\dots I}^c | N_{1\dots I}^c) \\ &\quad \times \frac{P(A_\alpha^c) \prod_{i=1}^I (\bar{\lambda}_i^c A_\alpha^c)^{N_i^c} e^{-\bar{\lambda}_i^c A_\alpha^c} / N_i^c!}{\int dA_\alpha^c P(A_\alpha^c) \prod_{i=1}^I (\bar{\lambda}_i^c A_\alpha^c)^{N_i^c} e^{-\bar{\lambda}_i^c A_\alpha^c} / N_i^c!}. \end{aligned} \quad (\text{B1})$$

Here, we have used Bayes theorem in the second step, and assumed conditional independence between A_α^c and $\bar{\lambda}_{1\dots I}^c$ in the third. We have also made use of equation (A1). Using the same Jeffreys prior as before, $P(A_\alpha^c) = \kappa_\alpha / A_\alpha^c$, and defining $N_{\text{tot},\alpha}^c = \sum_{i=1}^I N_i^c$ and $\bar{\lambda}_{\text{tot},\alpha}^c = \sum_{i=1}^I \bar{\lambda}_i^c$ we find

$$\begin{aligned} &P(\bar{\lambda}_{1\dots I}^c, A_\alpha^c | N_{1\dots I}^c) \\ &= P(\bar{\lambda}_{1\dots I}^c | N_{1\dots I}^c) \frac{1}{A_\alpha^c} \frac{(A_\alpha^c \bar{\lambda}_{\text{tot},\alpha}^c)^{N_{\text{tot},\alpha}^c} e^{-A_\alpha^c \bar{\lambda}_{\text{tot},\alpha}^c}}{\Gamma(N_{\text{tot},\alpha}^c)}. \end{aligned} \quad (\text{B2})$$

The BORG Markov chain provides samples from the posterior distribution $P(\bar{\lambda}_{1\dots I}^c | N_{1\dots I}^c)$. By marginalizing over $\bar{\lambda}_{1\dots I}^c$ using Monte-Carlo integration, we can estimate the posterior distribution for A_α^c :

$$\begin{aligned} &P(A_\alpha^c | N_{1\dots I}^c) \\ &\approx \frac{(A_\alpha^c)^{N_{\text{tot},\alpha}^c - 1}}{(N_{\text{tot},\alpha}^c - 1)!} \frac{1}{S} \sum_{s=1}^S (\bar{\lambda}_{\text{tot},\alpha,s}^c)^{N_{\text{tot},\alpha}^c} e^{-A_\alpha^c \bar{\lambda}_{\text{tot},\alpha,s}^c}. \end{aligned} \quad (\text{B3})$$

Here, we have taken S samples from the posterior, and $\bar{\lambda}_{\text{tot},\alpha,s}^c$ is calculated from the MCMC sample s . We therefore obtain the

expectation value of A_α^c as

$$E(A_\alpha^c | N_{1\dots I}^c) = \int_0^\infty dA_\alpha^c A_\alpha^c P(A_\alpha^c | N_{1\dots I}^c) \\ \approx \frac{N_{\text{tot},\alpha}^c}{S} \sum_{s=1}^S \frac{1}{\bar{\lambda}_{\text{tot},\alpha,s}^c}, \quad (\text{B4})$$

yielding equation (4). An analogous computation gives the expectation value for the mean number of galaxies in pixel i , $A_\alpha^c \bar{\lambda}_i^c$, yielding equation (5):

$$E(A_\alpha^c \bar{\lambda}_i | N_{1\dots I}^c) \approx \frac{N_{\text{tot},\alpha}^c}{S} \sum_{s=1}^S \frac{\bar{\lambda}_{i,s}^c}{\bar{\lambda}_{\text{tot},\alpha,s}^c}. \quad (\text{B5})$$

APPENDIX C: POSTERIOR PREDICTIVE TESTS FOR OTHER CLUSTERS

We show in Fig. C1, the posterior predictive tests for the remaining seven clusters considered in this work. For all these clusters, the posterior predictive tests pass in all but a handful of radial bins.

This paper has been typeset from a TeX/LaTeX file prepared by the author.

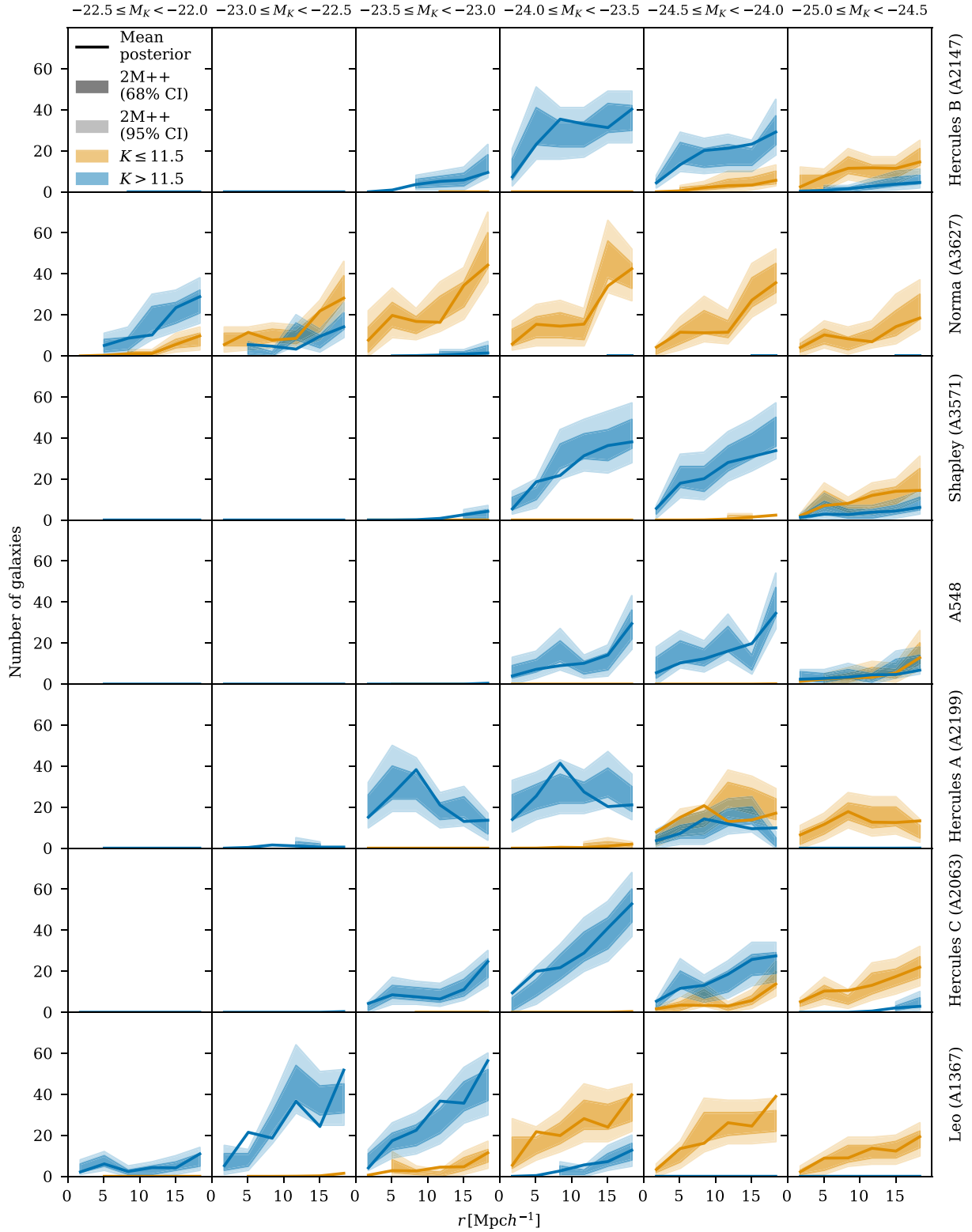


Figure C1. Posterior predictive tests for the other seven clusters considered in Fig. 4, using the same conventions as Fig. 5.