Statistics of extreme objects in the Juropa Hubble Volume simulation

William A. Watson,1† Ilian T. Iliev,1 Jose M. Diego,2 Stefan Gottlöber,3 Alexander Knebe,4 Enrique Martínez-González2 and Gustavo Yepes4

1Astronomy Centre, Department of Physics & Astronomy, Pevensey II Building, University of Sussex, Falmer, Brighton BN1 9QH, UK
2IFCA, Instituto de Física de Cantabria (UC-CSIC), Avda. Los Castros s/n, E-39005 Santander, Spain
3Leibniz-Institute for Astrophysics, An der Sternwarte 16, D-14482 Potsdam, Germany
4Universidad Autónoma de Madrid, Grupo de Astrofísica, E-28049 Madrid, Spain

Accepted 2013 November 7. Received 2013 November 6; in original form 2013 May 8

ABSTRACT

We present the first results from the Juropa Hubble Volume project, based a large N-body, dark matter-only cosmological simulation with a volume of $V = (6 h^{-1} \text{Gpc})^3$, containing 6000$^3$ particles, performed within the concordance $\Lambda$ cold dark matter cosmological model. The simulation volume is sufficient to probe extremely large length-scales in the universe, whilst at the same time the particle count is high enough so that dark matter haloes down to $1.5 \times 10^{12} h^{-1} \text{M}_\odot$ can be resolved. At $z = 0$, we identify over 400 million haloes. The cluster mass function is derived using three different halofinders and compared to fitting functions in the literature. The distribution of clusters of maximal mass across redshifts agrees well with predicted masses of extreme objects, and we explicitly confirm that the Poisson distribution is very good at describing the distribution of rare clusters. The Poisson distribution also matches well the level to which cosmic variance can be expected to affect number counts of high-mass clusters. We find that objects like the Bullet cluster exist in the far-tail of the distribution of mergers in terms of relative collisional speed. We also derive the number counts of voids in the simulation box for $z = 0, 0.5$ and 1.

Key words: methods: numerical – galaxies: haloes – large-scale structure of Universe.

1 INTRODUCTION

Surveys mapping a substantial portion of the observable Universe (e.g. BOSS – Dawson et al. 2013, WiggleZ – Drinkwater et al. 2010, BigBOSS – Schlegel et al. 2009, PanSTARRS – Magnier et al. 2013, DES – Mohr et al. 2012, PAU – Benítez et al. 2009, LSST – LSST Dark Energy Science Collaboration 2012, Euclid – Amiaux et al. 2012, etc.) aim to constrain the cosmological model to unprecedented accuracy. As they will be able to capture very faint objects they will have a shot-noise level low enough to be close to sampling variance limited. This requires impressive handling of every step of the observational pipeline in order to limit the possibility of systematic errors that may degrade the information contained in them. Aside from these observational efforts, there will also be a similarly high demand placed on our ability to generate theoretical predictions that are equally accurate. This undoubtedly calls for numerical simulations of cosmic structure formation that resolve galactic scales in volumes comparable to the ones covered by these surveys. This is a non-trivial task. A simulation must cover a wide dynamic range in order to accurately sample large-scale structure (LSS) in the universe. In particular, simulations need to resolve dark matter haloes, which are believed to host the observed galaxies, groups and clusters of galaxies, and accurately model the physics of galaxy formation and other non-linear physics, whilst adequately sampling large-scale matter fluctuations. Only recently have such simulations become feasible, and nowadays full-box simulations of considerable fractions of the observable Universe are being conducted utilizing close to a trillion particles (for a review of dark matter N-body simulations, see Kuhlen, Vogelsberger & Angulo 2012).

Whilst a careful comparison of the statistical clustering properties of objects, in particular galaxies, will put tighter constraints on the parameters of any cosmological model, it is worth noting that the mere existence of individual outliers might pose challenges. Following observations of a series of apparently extreme objects such as high-mass clusters (for example XMMU J2235.3−2557 – a cluster with mass $M > 4 \times 10^{14} h^{-1} \text{M}_\odot$ at redshift $z \sim 1.4$; Mullis et al. 2005; Rosati et al. 2009) or high-velocity collisions (for example the Bullet cluster; see Section 3.4), some authors have claimed that such objects are highly unlikely to exist in a concordance $\Lambda$ cold dark matter ($\Lambda$CDM) cosmology and hence pose a challenge to its validity (Jimenez & Verde 2009; Lee & Komatsu 2010; Cayón, Gordon & Silk 2011; Enqvist, Hotchkiss & Taanila 2011; Hoyle, Jimenez & Verde 2011; Holz & Perlmutter 2012). However,
others have found no tension in their analyses (Harrison & Coles 2011; Mortonson, Hu & Huterer 2011; Waizmann et al. 2012b) and it has been recently pointed out by Hotchkiss (2011) that many studies into high-mass cluster observations have used biased statistical methods, a result that has been corroborated by other studies (Hoyle et al. 2012; Waizmann, Ettori & Moscardini 2012a; Stalder et al. 2013). The case of the Bullet cluster is less clear and we discuss it in Section 3.4. Debate of this nature highlights the need to clarify our understanding of the statistics of such rare objects, for example by using large cosmological simulations. But this again requires simulations of large enough volumes and sufficient resolution to properly capture the likelihood of the formation of such rare clusters. Theoretical (as opposed to numerical) studies of such objects are challenging too due to their rarity and highly non-linear nature.

In this work, we present one of the largest cosmological dark matter-only simulations to date, the so-called Europa Hubble Volume (Jubilee) Universe, consisting of 6000^3 particles in a cubical volume of side-length 6 h⁻¹ Gpc. In this paper, we focus on a presentation of the simulation itself and its general properties with respect to the cosmic web, clustering properties and halo statistics. Subsequent papers will focus on topics including the generation of mock catalogues of luminous red galaxies (LRGs) and quasars, a calculation of the integrated Sachs–Wolfe (ISW) effect signal and its cross-correlation to LSS, the weak lensing signal and the SZ effect (see Watson et al. 2013b for initial results in some of these areas). This paper is laid out as follows. In Section 2, we outline our methodology for running the simulation and deriving from it results including halo and void catalogues. In Section 3, we present our main results and in Section 4 briefly discuss their potential implications.

2 METHODS

2.1 Simulations

The results presented in this work are based on two LSS N-body simulations, whose parameters are listed in Table 1. Our main simulation has 6000^3 (216 billion) particles in a volume of 6 h⁻¹ Gpc. The particle mass is 7.49 × 10¹¹ h⁻¹ M⊙, yielding a minimum resolved halo mass (with 20 particles) of 1.49 × 10¹² h⁻¹ M⊙, corresponding to galaxies slightly more massive than the Milky Way. LRGs (M ~ 10¹³ h⁻¹ M⊙) are resolved with 100 particles, and galaxy clusters (M > 10¹⁴ h⁻¹ M⊙) are resolved with 10³ particles or more. This main simulation is accompanied by a second, smaller ‘control’ one with 3072^3 (29 billion) particles in a volume of 3.072 h⁻¹ Gpc and exactly the same minimum resolution. We used the CUBEPM N-body code, a P³M (particle–particle–particle–mesh) code (Harnois-Deraps et al. 2012). It calculates the long-range gravity forces on a two-level mesh and short-range forces exactly by direct summation over local particles. The code is massively parallel, using hybrid (combining MPI and OpenMP) parallelization and has been shown to scale well up to tens of thousands of computing cores (see Harnois-Deraps et al. 2012 for complete code description and tests). Both simulations and most analyses were performed on the Europa supercomputer at the Jülich Supercomputing Centre in Germany (17 664 cores, 53 TB RAM, 207 TFlops peak performance) and required approximately 70 000 and 1.5 million core-hours for the 3 and 6 h⁻¹ Gpc boxes, respectively. The larger simulation was run on 8000 computing cores (1000 MPI processes, each with eight OpenMP threads) and the smaller one on 2048 cores.

2.1.1 Cosmology

We base our simulation on the 5-yr Wilkinson Microwave Anisotropy Probe (WMAP) results (Dunkley et al. 2009; Komatsu et al. 2009). The cosmology used was the ‘Union’ combination from Komatsu et al. (2009), based on results from WMAP, baryonic acoustic oscillations (BAO) and high-redshift supernovae, i.e. Ω_m = 0.27, Ω_λ = 0.73, h = 0.70, Ω_b = 0.044, σ_8 = 0.80 and n_s = 0.96. These parameters are similar to the recent cosmology results of the Planck Collaboration (Planck Collaboration et al. 2013), where, considering a combination of data from the Planck, WMAP and LSS surveys (showing baryon acoustic oscillations), the parameters were calculated to be Ω_m = 0.307 ± 0.0042, Ω_λ = 0.692 ± 0.010, h = 0.678 ± 0.0077, Ω_b = 0.048 ± 0.00052, σ_8 = 0.826 ± 0.012 and n_s = 0.9608 ± 0.00024. The power spectrum and transfer function used for setting initial conditions were generated using CAMB (Lewis, Challinor & Lasenby 2000). The initial condition generator employed in the run uses first-order Lagrangian perturbation theory, i.e. the Zel’dovich approximation (Zel’dovich 1970), to place particles in their starting positions. The initial redshift when this step takes place was z = 100. For a more detailed commentary on the choice of starting redshift for this simulation see Watson et al. (2013a).

2.2 Halofinding

We use two complementary definitions of haloes in this study. The first is the spherical overdensity (SO) definition of Lacey & Cole (1993). In this approach, haloes are taken to be spheres that have overdensities that are above a chosen threshold, Δ. The mass enclosed in these spheres is then given by

\[ M_Δ = \frac{4\pi\Delta \rho_m}{3} R_Δ^3, \]

where \( R_Δ \) is the radius of the halo and \( \rho_m \) is the background matter density in the Universe. We choose the overdensity threshold to be Δ/Δₗₗₗₗ, i.e. an overdensity of 178 times the background matter density. This is a common choice motivated from the top-hat model of non-linear collapse in an Einstein–de Sitter universe (Gunn & Gott 1972).

The second halo definition we adopt is that of the friends-of-friends (FOF) algorithm, first proposed by Davis et al. (1985). Haloes defined by this algorithm are identified within a simulation volume as agglomerations of particles that lie within a certain overdensity. This is a common choice from the ‘linking length’ × the mean interparticle separation of particles in the simulation. Groups of particles within this distance of each other are identified as individual dark matter haloes.

### Table 1. N-body simulation parameters. Background cosmology is based on the WMAP 5-yr results.

<table>
<thead>
<tr>
<th>Boxsize</th>
<th>N_part</th>
<th>Mesh</th>
<th>Smoothing</th>
<th>m_particle</th>
<th>M_halo.min</th>
</tr>
</thead>
<tbody>
<tr>
<td>3072 h⁻¹ Mpc</td>
<td>3072³</td>
<td>6144³</td>
<td>50 h⁻¹ kpc</td>
<td>7.49 × 10¹¹ h⁻¹ M⊙</td>
<td>1.49 × 10¹² h⁻¹ M⊙</td>
</tr>
<tr>
<td>6000 h⁻¹ Mpc</td>
<td>6000⁶</td>
<td>12 000³</td>
<td>50 h⁻¹ kpc</td>
<td>7.49 × 10¹¹ h⁻¹ M⊙</td>
<td>1.49 × 10¹² h⁻¹ M⊙</td>
</tr>
</tbody>
</table>
For our FOF haloes, we follow various previous authors (Jenkins et al. 2001; Reed et al. 2003, 2007; Crocce et al. 2010; Courtin et al. 2011; Angulo et al. 2012) and use a linking length of 0.2. For further analysis on how the choice of halofinding parameters affect the mass function see Watson et al. (2013a) and references therein.

We employ three halofinding codes in our analysis: CUBEP3M’s own on-the-fly SO halofinder (hereafter ‘CPMSO’; Harnois-Deraps et al. 2012), the Amiga Halo Finder (hereafter ‘AHF’; Gill, Knebe & Gibson 2004; Knollmann & Knebe 2009) and the FOF halofinder from the GADGET-3 N-body cosmological code (Springel 2005, an update to the publicly available GADGET-2 code).

The CPMSO halofinder utilizes a fine mesh from the CUBEP3M code (with spacing twice as fine as the mean interparticle separation) and an interpolation scheme to identify local peaks in the density field. The code first builds the fine-mesh density using either cloud-in-cell (CIC) or nearest-grid-point interpolation. It then proceeds to search for and record all local density maxima above a certain threshold (typically set to 100 above the mean density) within the physical volume. It then uses quadratic interpolation on the density field to determine more precisely the location of the maximum within the densest cell. The halo centre determined this way agrees closely with the centre of mass of the halo particles. Each of the halo candidates is inspected independently, starting with the highest peak. The grid mass is accumulated in spherical shells of fine grid cells surrounding the maximum until the mean overdensity within the halo drops below \( \Delta \). While the mass is accumulated it is removed from the mesh, so that no mass element is double-counted. For further details on the CPMSO method see Harnois-Deraps et al. (2012).

The halofinder AHF\(^1\) is a SO finder that identifies (isolated and sub) haloes as described in Gill, Knebe & Gibson (2004) and Knollmann & Knebe (2009). It employs a recursively refined grid to locate local overdensities in the density field. The identified density peaks are then treated as centres of prospective haloes. The resulting grid hierarchy is further utilized to generate a halo tree containing the information of which halo is a (prospective) host and subhalo, respectively. Halo properties are calculated based on the list of particles asserted to be gravitationally bound to the respective density peak. For a comparison of its performance to other finders in the field, we refer the reader to Knebe et al. (2011, 2013) and Onions et al. (2012).

The specifics of the FOF halofinder packaged in with the GADGET-3 code currently have not been detailed in any publication but the algorithm itself is outlined in Davis et al. (1985). The main difference in the algorithm that exists in the GADGET-3 version is that the code is parallelized for distributed-memory machines. Specifically, haloes are found in local subvolumes of the simulation assigned to individual MPI tasks (created using the GADGET-3 domain decomposition which utilizes a space-filling Peano–Hilbert curve – for details see the GADGET-2 paper, Springel 2005) and then haloes that extend spatially beyond the edges of the subvolumes are linked together in a final MPI communication step. We have altered the GADGET-3 code to read CUBEP3M’s particle output format and significantly reduced its memory footprint by stripping away extraneous data structures.

Due to limitations in the scaling of the codes with processor numbers and the large memory footprint of the Jubilee simulation, it was necessary to split the simulation time-slices into 27 subvolumes and run the halofinding algorithms on each subvolume independently.

\(^1\) AHF is freely available from http://www.popia.ft.uam.es/AMIGA.

Each subvolume included a buffer zone which overlapped with the neighbouring ones, for correct handling of haloes straddling two or more subregions. We then stitched the subvolumes back together to create the final AHF and FOF halo catalogues, removing any duplicated structures in the overlapping buffers. This approach allows the handling of much larger amounts of data than otherwise possible and provides additional flexibility in terms of computational resources needed for post-processing.

### 2.3 Void finding

The formation of structure in the Universe is a hierarchical process: small objects form, grow by accretion and merging, and form more and more massive objects up to clusters of galaxies. Between the clusters large filaments can be seen both in observational data as well as in numerical simulations (Fig. 1). These filaments surround large regions of low density which do not contain objects as massive as the ones found in the filaments or the knots at the end of filaments. These low-density regions – voids – are the most extended objects in the Universe. There are many different ways to define voids and correspondingly there are many different void-finding algorithms (for a review, see Colberg et al. 2008). In the following, we are interested in the largest spherical regions of the Universe which do not contain any object above a certain threshold in mass. In principle, one could extend this definition of voids also to non-spherical regions; however, in this case one can get arbitrary volumes depending on the shape allowed. Since we are interested in the void function, we restrict ourselves to spherical voids which are described only by one parameter, their radius. We identify voids in a sample of point-like objects distributed in space. Here, these objects are our AHF haloes above a certain mass, but one could also use galaxies above a certain luminosity. Thus, our voids are characterized by a threshold mass. If one decreases this mass threshold, the number of objects increases and the size of the void decreases. In fact, a given void defined with objects at a higher mass becomes decomposed into many smaller voids defined in the distribution of lower mass objects (Gottlöber et al. 2003). This reflects the scale-free nature of structure formation. The algorithm searches first for the largest empty sphere, then repeats taking into account the previously found voids so that no region is double-counted.

### 2.4 Online data bases

It is our intention to make the data from the Jubilee simulation publicly available. These data will consist of three complementary halo catalogues of CPMSO, AHF and FOF haloes, in addition to LRG catalogues derived from the halo data and a catalogue of voids. The CPMSO will be available across a wide number of redshifts (~30) from \( z = 0–6 \), whereas the AHF and FOF data will be initially available only for \( z < 1 \). Further, data sets will include smoothed density fields and maps of weak lensing and ISW signals. An SQL data base has been set up so that the data can be queried to suit the requirements of individual users. Further information can be found at the Jubilee project website: http://jubilee-project.org.

### 3 RESULTS

#### 3.1 Large-scale structure and the cosmic web

In Fig. 1, we show a slice of the cosmic web at \( z = 0 \) extracted from our 6 \( h^{-1} \) Gpc simulation. Perhaps, most striking is the homogeneity of the matter distribution at large scales. This is expected from the
cosmological principle which states that, on large enough scales, the Universe is homogeneous and isotropic. On smaller scales, significant features in the density field can be observed including voids, walls, filaments and clusters. For any simulated observations (for example of ISW and weak lensing signals; Watson et al. 2013b), we place virtual observers inside the simulation volume at a given location. As can be seen in Fig. 1, the full sky as observed by an observer will show a highly homogeneous distribution of galaxies past a proper distance of a few hundred Mpc (i.e. a redshift of around $z \sim 0.1$).

In Fig. 2, we show the evolution of the power spectra of the density field, $P(k)$, from redshift $z = 6$ to 0. The particles were interpolated on to a regular grid of 12 000$^3$ cells using the CIC interpolation scheme. From these data we then applied a correction for aliasing and the CIC window function and another for the effect of Poisson noise, all based on the prescription laid out in Jing (2005).

The BAO scale, $k \sim 0.1 \, h \, \text{Mpc}^{-1}$, is well within the simulation box size and the BAO are visible in the power spectra. At high redshift, $z \sim 6$, the power spectrum is largely linear, except at the smallest scales ($k > 1 \, h \, \text{Mpc}^{-1}$), where the power grows faster than the linear growth factor predicts. As the hierarchical structure formation proceeds, this non-linearity scale propagates to ever larger scales, reaching $k \sim 0.1 \, h \, \text{Mpc}^{-1}$ at redshift $z = 0$, and thereby affecting the BAO scale.

We calculate halo mass functions using the three halofinding algorithms outlined in Section 2.2. Fig. 3 shows the residuals between our haloes and two fits from the literature at $z = 0$. We compare our CPMSO and AHF haloes to the Tinker et al. (2008) mass function, noting that the Tinker et al. (2008) fit was calibrated to haloes with an overdensity criteria of $\Delta = 200$ versus the background matter density and our haloes were calculated using a value of $\Delta = 178$. Despite this, we see a good correspondence to within $\sim$5 per cent between the Tinker et al. (2008) fit and our AHF data for haloes with particle counts greater than $\sim$300. For the very largest haloes, there is evidence that the Tinker et al. (2008) fit may be overpredicting the mass function, although this is where shot noise begins to severely affect number counts of objects. The CPMSO data follow a similar trend to the AHF data. We overlay on these plots two of the fits from Watson et al. (2013a) – mass function results calibrated to data that included the Jubilee haloes presented here. The fit used for the left-hand panel is the redshift-dependent fit based on the CPMSO halofinder, the fit for the central panel is a
3.2 Cosmic variance

Due to the large size of our simulated volume, we are able to quantify cosmic variance on scales smaller than our box size in terms of the number counts of objects one expects to find in a given volume. To that end, we have compared halo counts in different mass bins in different sized subvolumes. We chose the subvolumes such that they filled the entire full box with no overlap. The results are shown in Fig. 4. We show the 1 standard deviation error in the number counts of haloes by mass bin relative to our entire \((6h^{-1}\text{Gpc})^3\) volume for subbox lengths of 3, 2, 1 and 0.5 \(h^{-1}\text{Gpc}\). These choices directly compare to the box lengths of some contemporary simulations (Millennium-XXL – Angulo et al. 2012, Horizon – Teyssier et al. 2009, MultiDark – Prada et al. 2012 and Millennium – Springel et al. 2005, respectively). We also show a prediction for this error calculated by assuming that the halo number counts follow the CPMSO redshift parametrized mass function from Watson et al. (2013a) and by assuming that the observed error in number counts follows a Poisson distribution. This theoretical prediction matches the high-mass data very well. For lower masses, the error becomes dominated by sample variance, as discussed in Smith & Marian (2011), and the Poisson prediction presented here begins to break down. We discuss how well the Poisson distribution matches the counts of high-mass clusters in Section 3.3 below.

As expected, the error is minimal for lower mass haloes and increases for rarer objects. At \(z = 0\), the 0.5 \(h^{-1}\text{Gpc}\) box has an error of under 10 per cent up until haloes of mass around \(4 \times 10^{14}h^{-1}\text{M}_\odot\), while for box lengths of 1, 2 and 3 \(h^{-1}\text{Gpc}\), a 10 per cent error in number counts per mass interval is realized at around \(1 \times 10^{15}, 2 \times 10^{15}\) and \(3 \times 10^{15}h^{-1}\text{M}_\odot\), respectively. The errors are exacerbated at higher redshifts, due to the haloes of a fixed size growing rarer at earlier times.

One subtlety should be mentioned: because the subvolumes considered in this analysis were derived from a larger simulation volume, they include the effect of matter fluctuations that exist on scales larger than their box lengths. We stress here that this is not the case for simulations with equivalent volumes to these subvolumes, as modes of power in the density field that are larger than the box length of a simulation are typically set to zero. This implies that the variation in number counts presented here is slightly
Extreme objects in Jubilee

Figure 4. Relative error in $dn/dM$ for subvolumes of box lengths 3, 2, 1 and 0.5 $h^{-1}$ Gpc. ‘Relative error’ is defined as the standard deviation from the expected number counts per unit volume in a given mass interval, as defined using the entire $6h^{-1}$ Gpc volume. The solid lines are the predictions based on the AHF redshift parametrized mass function from Watson et al. (2013a) combined with the assumption that the error in number counts around the mean is given by the Poisson distribution.

3.3 Statistics of rare objects

A current topic in cosmology that relates to the number counts of very high mass objects is that whether large observed clusters are in conflict with the standard $\Lambda$CDM model. In Fig. 5, we show a theoretical prediction for the expected distribution of maximal mass clusters. This prediction was created using the extreme value statistics (EVS) prescription of Harrison & Coles (2011) and the redshift parametrized redshift-dependent mass function, based on CPMSO haloes, from Watson et al. (2013a). It is comparable to the plot in fig. 1 of Harrison & Coles (2012) which was created using the mass function from Tinker et al. (2008), except for the fact that the mass of the haloes in this version is taken to be set by the $\Delta = 178$ overdensity criterion, rather than the $\Delta = 200$ criterion used in Harrison & Coles (2012). The black data points shown in Fig. 5 correspond to the largest clusters observed in the Jubilee simulation by a central observer, also based on a $\Delta = 178$, as per the configuration of the CPMSO halofinder. The redshift shells for both the EVS contours and the maximal mass clusters are identical. The data all lie within the 3$\sigma$ range showing the expected result that there is no tension between objects observed in a $\Lambda$CDM cosmological simulation and the theoretical expectation from EVS.

Of interest is whether there are observed clusters in the Universe that have masses that are in tension with the $\Lambda$CDM model. To date, observations have shown this not to be the case, as shown in a systematic review by Harrison & Hotchkiss (2013).
of the Poisson distribution in cluster counts across different masses
see Smith & Marian 2011, who used simulations of box length 1.5 $h^{-1}$ Gpc for their study).

3.4 High $\Delta v$ mergers and the Bullet cluster
There has been recent debate regarding whether the Bullet cluster
(1E0657-56, which resides at a redshift of $z = 0.296$) poses a challenge
to the $\Lambda$CDM model. 1E0657-56 consists of a large cluster
of mass $M_{500} \sim 1.5 \times 10^{15}$ $M_{\odot}$ and a subcluster – the ‘bullet’
– of mass $M_{500} \sim 1.5 \times 10^{14}$ $M_{\odot}$ that has traversed through
the larger cluster, creating a substantial bow shock along the way
(Barrena et al. 2002; Markevitch et al. 2002; Clowe, Gonzalez &
Markevitch 2004; Bradač et al. 2006; Clowe et al. 2006). Tension
with $\Lambda$CDM arises from the calculated value for the speed of
the shock of $v_s = 4740 \pm 710$ km s$^{-1}$ (Markevitch et al. 2002, 2004;
Springel & Farrar 2007), which was originally calculated to be
too high for a $\Lambda$CDM universe (Farrar & Rosen 2007) – whereas
it might be better accommodated in alternative cosmologies (e.g.
Llinares, Zhao & Knebe 2009). Other studies have concluded that
the velocity is not in tension with $\Lambda$CDM (Hayashi & White 2006).
An important clarification of this issue was presented by groups
working on simulations of Bullet-like systems (Takizawa 2005, 2006;
Milosavljević et al. 2007; Springel & Farrar 2007; Mastropietro & Burkert 2008) where, in general, it was found that the shock
speed was substantially higher than the speed of the mass centroid
of the infalling subcluster. For example, Springel & Farrar (2007)
found that a Bullet-like system in their simulations had a shock speed of $\sim 4500$ km s$^{-1}$ whereas the subcluster had a speed of only
$\sim 2600$ km s$^{-1}$. Milosavljević et al. (2007) found that in an illustrative
simulation, the subcluster CDM halo had a speed that was 16 per cent lower than that of the shock.
Even given this moderation of the extreme subcluster speed in
1E0657-56 there have still been claims in the literature that the $\Lambda$CDM model may be incapable of creating such a system (Lee &
Komatsu 2010; Thompson & Nagamine 2012). This is not wholly
unexpected as (a) Mastropietro & Burkert (2008) have shown that
the properties of the bow shock are not well described by simulations
and (b) even with a moderation in subcluster speed along the lines
of Springel & Farrar (2007) or Milosavljević et al. (2007), the
speed may still be too high for the $\Lambda$CDM model to accommodate.
These studies have relied on numerical simulations to observe the
distribution of relative velocities in colliding clusters. From these
distributions, 1E0657-56 can be assessed and deemed to be either
rare for a $\Lambda$CDM universe or so rare that it puts the whole model in
doubt.

Alternative approaches have also been taken in addressing this
question. Forero-Romero, Gottlöber & Yepes (2010) looked in 2D-
projected position-space for Bullet-like systems in the MareNostrum Universe, a large hydrodynamical cosmological simulation.
The characteristic distribution of gas and dark matter in 1E0657-
56, as projected on the sky – with a large displacement between
the cluster’s gas and dark matter – was found to be expected in 1–
2 per cent of clusters with masses larger than $10^{14}$ $h^{-1}$ $M_{\odot}$. Nusser
(2008) performed a ‘back in time’ analysis to place bounds on the
relative overdensity the system resides in, in the Universe, conclud-
ing that for a relative speed of $\sim 4500$ km s$^{-1}$, the system would
need to have a mass of $2.8 \times 10^{15}$ $h^{-1}$ $M_{\odot}$ and exist in a local
overdensity of 10 times the background density of the Universe.

Here, we use the huge number counts of clusters in the Jubilee
simulation to add to the debate. We consider AHE (sub) haloes with
mass greater than $1 \times 10^{15}$ $h^{-1}$ $M_{\odot}$ that are colliding with (host)
Extreme objects in Jubilee

Figure 7. Relative pairwise velocities for haloes. The data points correspond to the observed Bullet cluster speed from Markevitch et al. (2004) (cross) and the corrected speed estimated by Springel & Farrar (2007) (circle). The simulated speeds were obtained from the full box using the output redshift slice ($z = 0.32$) that best matched the redshift of the Bullet cluster ($z = 0.296$). The halofinder used was AHF, which can resolve both subhaloes and haloes. We have shown in blue, merging objects that are both host haloes in their own rights and in red, merging objects where a subhalo is moving with a high relative velocity to its host halo. The data in the simulation were reduced by considering haloes with a mass of over $1 \times 10^{13} h^{-1} M_\odot$ and data points were plotted with the restriction that one of the haloes in the pair had a mass of at least $7 \times 10^{14} h^{-1} M_\odot$. Finally, we have added in a random scatter to mimic the effect of Eddington bias, as described in the text.

Our results are shown in Fig. 7, along with the original Bullet speed presented in Markevitch et al. (2004), and the moderated result from Springel & Farrar (2007), which represents the lowest value from the literature to date. We show in blue, haloes that are colliding pairs and in red, haloes that are a colliding subhalo and halo pair. We have added a normally distributed random scatter to our velocities with a width given by the error in the observed value for $v_s$. This is to mimic the effect of Eddington bias in our simulated data. This observational bias arises when observing extreme measurements in a distribution of measurements all with some associated scatter. As there are many more data points that exist with less extreme velocities than the one in question, it is likely that an extreme data point is an upscattered less extreme one. As we know very precisely the pairwise velocities of halo pairs in our simulation, adding random scatter to this distribution serves to create the effect in our measurement.

As can be seen from the distribution, the Bullet cluster is an extreme object, but only when the radial separation of the halo pairs is considered. We find many candidate mergers in our volume with a collision speed that equals or exceeds the more conservative speed estimate for the cluster and a few objects that are not far from the higher velocity estimate of Markevitch et al. (2004). However, we find no objects that, at a closer separation, give rise to a large enough merging velocity. This is likely to be due to the effect of only considering one simulation output in our analysis. At any given output redshift only a handful of haloes will be undergoing a major merger event of the sort we are interested in and this is reflected by the paucity of data points that lie at a separation of less than $0.6 h^{-1} \text{Mpc}$. It is likely, therefore, that over the course of the Jubilee simulation run, high-velocity mergers of the type observed in the Bullet cluster do occur.

This result is in line with previous attempts to use large cosmological simulations to address this issue where the bullet was not found to be extreme (Hayashi & White 2006; Thompson & Nagamine 2012). Interestingly, Thompson & Nagamine (2012) extrapolated their results from smaller simulation volumes and concluded that a volume of $(4.5 h^{-1} \text{Gpc})^3$ would be required in order to observe a Bullet-like cluster.

The conclusion that we put forward based on this result is that there is at present no tension between our data and the standard cosmological model. This conclusion is tentative, however, and there would appear to be a need for careful further research into this question based on a number of points. First, the results are very sensitive to the mass cuts imposed on the candidate search. As we now very precisely the pairwise velocities of halo pairs in our simulation, adding random scatter to this distribution serves to create the effect in our measurement.

As can be seen from the distribution, the Bullet cluster is an extreme object, but only when the radial separation of the halo pairs is considered. We find many candidate mergers in our volume with a collision speed that equals or exceeds the more conservative speed estimate for the cluster and a few objects that are not far from the higher velocity estimate of Markevitch et al. (2004). However, we find no objects that, at a closer separation, give rise to a large enough merging velocity. This is likely to be due to the effect of only considering one simulation output in our analysis. At any given output redshift only a handful of haloes will be undergoing a major merger event of the sort we are interested in and this is reflected by the paucity of data points that lie at a separation of less than $0.6 h^{-1} \text{Mpc}$. It is likely, therefore, that over the course of the Jubilee simulation run, high-velocity mergers of the type observed in the Bullet cluster do occur.

This result is in line with previous attempts to use large cosmological simulations to address this issue where the bullet was not found to be extreme (Hayashi & White 2006; Thompson & Nagamine 2012). Interestingly, Thompson & Nagamine (2012) extrapolated their results from smaller simulation volumes and concluded that a volume of $(4.5 h^{-1} \text{Gpc})^3$ would be required in order to observe a Bullet-like cluster.

The conclusion that we put forward based on this result is that there is at present no tension between our data and the standard cosmological model. This conclusion is tentative, however, and there would appear to be a need for careful further research into this question based on a number of points. First, the results are very sensitive to the mass cuts imposed on the candidate search. It would be very difficult to find a precise analogue to the Bullet cluster in terms of the masses, velocities and spatial separation of the haloes. Here, we have taken a cut-off in mass that allows us to search for Bullet-like systems rather than a precise Bullet cluster. Secondly, we have placed no restrictions on the directions of the relative velocities of the halo pairs. The bow shock observed in the Bullet cluster has arisen from the Bullet subhalo having passed through the parent halo (it is this occurrence, which fortuitously lies almost in the plane of the sky as we observe it, that has allowed us
to identify the relative pairwise velocities of the halo and subhalo in the system). In our analysis, we plot all the pairwise velocities of the haloes, making no distinction between haloes that are inflating and haloes that have already undergone a collision and not considering how the orientations of the collisions might appear to a specific observer. This is a fair way to assess the data as the actual collision in a Bullet cluster-like system is expected to take a few hundred megayears so is a relatively short event. Canvasing all our haloes in this manner assesses whether there is likely to be or whether there has been a Bullet cluster-like collision in the simulation around $z \sim 0.3$. Further studies should include how random observers would observe these events. Lastly, halo-finding algorithms are notoriously suboptimal when trying to find and separate haloes that are merging (this is discussed in detail in Knebe et al. 2011, we draw the reader’s attention to fig. 10 from that paper in particular). However, as the separation in question of the two haloes is relatively large, this is unlikely to be affecting our results.

3.5 The Jubilee void function

The distribution of voids for a given threshold is characterized by the void function, the number of spheres with radii larger than $R_{\text{void}}$ per volume. We have studied the void distribution at redshifts $z = 0, 0.5$ and 1. At $z = 0$, we have identified the voids in the distribution of haloes more massive than $5 \times 10^{14}, 2 \times 10^{14}, 1 \times 10^{14}$ and $1 \times 10^{13} h^{-1} M_\odot$. At redshift $z = 0$, we identified 244 989, 1753 982, 5596 627 and 91 615 821 haloes more massive than these thresholds, respectively. Thus, the mean distance between them (i.e. the box length divided by the cube root of the number counts) is about 96, 50, 34 and 13 $h^{-1}$ Mpc. Nevertheless, we found huge volumes which do not contain any of these objects.

In Fig. 8, top panel, we show the void functions at $z = 0$ for four different threshold masses. For the largest threshold, we find a few very large spheres with radii of $150 h^{-1}$ Mpc which do not contain any cluster more massive than $5 \times 10^{14} h^{-1} M_\odot$. For smaller thresholds, the void function is very steep, i.e. there are a number of voids with a volume almost as large as the volume of the largest voids defined by the threshold. This means that the voids are almost uniformly distributed, as there are so many of a similar size. At higher redshifts (middle and bottom panel of Fig. 8), we observe similar behaviour but, due to the evolution of the mass function, only with lower threshold masses. Note that at the lowest threshold ($10^{13} h^{-1} M_\odot$), the maximum void radius is almost redshift independent between $z = 0$ and occurs at a void radius of about $40 h^{-1}$ Mpc. This may seem in contradiction to the fact that low-density regions expand slightly faster than the mean expansion rate of the Universe. However, since the tracers of the voids are also evolving, the number of objects above the threshold evolves. For $10^{13} h^{-1} M_\odot$ mass haloes, the number counts rise from 38 994 056 at $z = 1$ to 91 615 821 at $z = 0$. Therefore, the mean distance shrinks from 18 to $13 h^{-1}$ Mpc, and using this threshold mass, we see the interesting result that the maximum void radius remains almost constant in time.

4 SUMMARY AND DISCUSSION

In this paper, we have presented a broad range of results from a ΛCDM-based simulation. The results have focused on predictions on very large scales, such as extremely massive clusters and large void regions. The simulation itself represents one of the largest undertaken to date, with a volume of $(6 h^{-1} \text{Gpc})^3$ and haloes resolved down to $1.4 \times 10^{12} h^{-1} M_\odot$, a resolution that allows the creation of mock LRG and cluster catalogues.

The distribution of dark matter haloes in the Jubilee was found to be well described by fitting functions from the literature, and the dark matter haloes from the Jubilee have been used in a separate paper (Watson et al. 2013a) to construct mass function fits across a broad range of redshifts and volumes. For the rare tail of the mass function, we have confirmed that the Poisson distribution describes well the number counts of objects. The masses of clusters with extremal masses in the Jubilee simulation were investigated...
across a range of redshifts and were found to agree well with both observation and theory, in particular the expected masses of the very largest objects found when using EVS.

4.1 Implications for precision cosmology

An important prediction from this simulation is the expected effect of cosmic variance on the counts of massive clusters. This result can be used to gauge number-count errors in survey and simulation data. Understanding this is a vital component of the drive towards high-precision cosmology. We showed in Fig. 4, how the expected number of clusters in given volumes are likely to vary. In general, quantifying the effect of cosmic variance in simulations is notoriously tricky due to the requirement for either multiple repeats of a simulation or for a large simulated volume (or preferably both of these). Due to the large scale of the Jubilee volume, we are able to use the latter and do so in a manner that includes the long-wavelength modes of the matter distribution. The variation in cluster counts for smaller boxes or surveys is highly significant if one is investigating the distribution of high-mass objects such as galaxy clusters, which form an important cosmological probe.

4.2 The largest voids

Our largest void, defined using a threshold mass of $5 \times 10^{14} h^{-1} M_\odot$, is $\sim 350 h^{-1}$ Mpc across. To put this void in context, it is around one fifth of the volume of the Millennium simulation and it contains no clusters with mass greater than $5 \times 10^{14} h^{-1} M_\odot$. The probability, based on volume-occupation alone, of finding yourself within this void in the universe represented by the Jubilee simulation is 0.01 per cent. There have been investigations as to whether our occupying a local underdensity might explain the apparent existence of an accelerated expansion in the late-time Universe (Ellis 1979; Mustapha, Hellaby & Ellis 1997; Zehavi et al. 1998; Tomita 2001; Iguchi, Nakamura & Nakao 2002; Barausse, Matarrese & Riotto 2005; Moffat 2005; Wiltshire 2005; Alexander et al. 2009; Febrary et al. 2010; Marra & Pääkkönen 2010; Nadathur & Sarkar 2011). The void in question would need to have very specific characteristics that include its radius, sphericity, density and density profile. Predictions for these void parameters vary but have typically required the void to be of at least a few hundred Mpc in radius and, importantly, close to spherically symmetric, with us as observers very near its centre. This latter requirement is due to the Type Ia supernovae data implying that dark energy is close to isotropic across the sky. We see from our void functions in Fig. 8 that there are a few hundred voids in the Jubilee volume with radii $R_{\text{void}} > 100 h^{-1}$ Mpc, for the $5 \times 10^{14} h^{-1} M_\odot$ mass threshold. We estimate the proportion of the entire simulation volume taken up by voids with a radius of $R_{\text{void}} > 100 h^{-1}$ Mpc to be 0.04 per cent. Adding an additional requirement that an observer occupies the central 1 per cent of the void volumes in question, we arrive at the total spatial volume in the Jubilee box that would contain observers in the centre of voids of radii greater than 100 $h^{-1}$ Mpc to be $\sim 0.0004$ per cent. This is a rough statistical estimate and ignores the fact that observers might be better considered to only exist at the locations of galaxies in the simulation. In addition, the simulation contains a dark energy component so it has already modelled the effect of late-time accelerated expansion on structure formation. This latter point does not alter the order of magnitude of the result as void sizes in universes without dark energy are comparable to void sizes in $\Lambda$CDM (Müller et al. 2000). We intend to look more closely into putting a probability on this figure for voids in the Jubilee simulation in a follow-up paper.

4.3 The $\Lambda$CDM model versus observations

The distribution of most massive clusters in the Jubilee was found to be in line with current theoretical predictions based on EVS. The nature of EVS is that it lacks predictive power in terms of constraining models, but it is a powerful method for ruling out models based on only a handful of extreme data points (various authors have previously implemented studies in cosmology based on it; for example Coles 1988; Antal et al. 2009; Colombi et al. 2011; Davis et al. 2011; Harrison & Coles 2011). Had the masses of observed clusters in Harrison & Hotchkiss (2013) lain significantly away from the expected EVS prediction, then the $\Lambda$CDM model would be immediately placed in doubt. One result in this paper that of the extreme nature of the Bullet cluster is suggestive of a possible tension with $\Lambda$CDM. An EVS approach is likely to cast this result in a more comprehensive context, but it is beyond the scope of this paper to attempt analysis along these lines.

In follow-up work, we intend to investigate in more detail the existing discrepancy between observations of the ISW signal and the expected $\Lambda$CDM signal. For example, Hunt & Sarkar (2010) claim that the observed voids from SDSS data are too large for a $\Lambda$CDM universe. This result was based largely on analysis of the ISW signal in Granett, Neyrinck & Szapudi (2008) and Granett, Neyrinck & Szapudi (2009). This highlights the intimate link between the void distribution and the ISW signal – underdense regions imprint themselves on the CMB via the ISW effect – and represents a current challenge to the $\Lambda$CDM model.

ACKNOWLEDGEMENTS

The simulation was performed on the Juropa supercomputer of the Jülich Supercomputing Centre (JSC). We would like to give special thanks to Alexander Schnurrfeil and the whole Jülich support team for their help during the phase of code implementation as well as during the production runs. We thank Leonidas Christodoulou, Peter Coles, Shaun Hotchkiss and Aurel Schneider for their helpful comments and feedback on this work. We also thank Steffen Knollmann for his help in applying the AHF halofinder to our data. We thank Volker Springel for allowing us to use the GADGET-3 code with its FOF halofinder. WW thanks the Southeast Physics Network (SEP-Net) for providing funding for his research. ITI was supported by the SEPNet and the Science and Technology Facilities Council grants ST/F002858/1 and ST/I000976/1. AK was supported by the Spanish Ministerio de Ciencia e Innovación (MICINN) in Spain through the Ramón y Cajal programme as well as the grants AYA 2009-13875-C03-02, AYA2009-12792-C03-03, CSD2009-00064, CAM S2009/ESP-1496 and the Ministerio de Economía y Competitividad (MINECO) through grant AYA2012-31101. GY acknowledges support from MINECO (Spain) under research grants AYA2009-13875-C03-02, FPA2009-08958, AYA2012-31101 and Consolider Ingenio SyeC CSD2007-0050. He also thanks Comunidad de Madrid for support under PRICIT project ASTROMADRID (S2009/ESP-146) and EMG acknowledges support of the consolider project CSD2010-00064 funded by the Ministerio de Economía y Competitividad and AYA2012-39475-C02-01.

REFERENCES
