Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **530,** 1129–1143 (2024) https://doi.org/10.1093/mnras/stae893
Advance Access publication 2024 March 30

Clustering of binary black hole mergers: a detailed analysis of the EAGLE + MOBSE simulation

Matteo Peron [1,2★] Andrea Ravenni,[2,3] Sarah Libanore,[2,3,4] Michele Liguori [2,3,5,6] and Maria Celeste Artale [2,3,7]

[1]*Dipartimento di Scienze Matematiche, Fisiche and Informatiche, Università di Parma, Parco Area delle Scienze 7/A, I-43124 Parma, Italy*
[2]*Dipartimento di Fisica e Astronomia G. Galilei, Università degli Studi di Padova Via Marzolo 8, I-35131 Padova, Italy*
[3]*INFN, Sezione di Padova, Via Marzolo 8, I-35131 Padova, Italy*
[4]*Department of Physics, Ben-Gurion University of the Negev, Be'er Sheva 84105, Israel*
[5]*Dipartimento di Fisica, Università degli Studi di Trento, Via Sommarive 14, I-38123 Povo (TN), Italy*
[6]*INAF - Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[7]*Departamento de Ciencias Físicas, Universidad Andrés Bello Fernandez Concha, 700 Las Condes, Santiago, Chile*

Accepted 2024 March 25. Received 2024 March 25; in original form 2023 July 12

ABSTRACT

We perform a detailed study of the cosmological bias of gravitational wave (GW) events produced by stellar binary black hole mergers (BBHM). We start from a BBHM distribution modelled inside the EAGLE hydrodynamical simulation using the population synthesis code MOBSE. We then compare our findings with predictions from different halo occupation distribution (HOD) prescriptions and find overall agreement, provided that the modelled properties of host galaxies and haloes in the semi-analytical treatment match those in the simulations. By highlighting the sources of these discrepancies, we provide the stepping stone to build future more robust models that prevent the shortcoming of both simulation-based and analytical models. Finally, we train a neural network to build a simulation-based HOD and perform feature importance analysis to gain intuition on which host halo/galaxy parameters are the most relevant in determining the actual distribution and power spectrum of BBHM. We find that the distribution of BBHM in a galaxy does not only depend on its size, star formation rate and metallicity, but also by its kinetic state, namely its total energy and velocity dispersion.

Key words: gravitational waves – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Future gravitational wave (GW) interferometers, such as the Einstein Telescope (ET, Punturo et al. 2010; Sathyaprakash et al. 2012; Maggiore et al. 2020) and Cosmic Explorer (CE, Evans et al. 2021), are expected to detect $\sim 10^4$–10^5 merger events per year; progenitors of such events will mainly be stars of masses of $\mathcal{O}(10-100)\,M_\odot$. In this regime, it will be possible to study the spatial distribution and clustering properties of GW sources in similar fashion as done for galaxies. These GW catalogues will contain a significantly smaller number of objects with respect to galaxy surveys and will have relatively poor sky localization, making the smallest scales unaccessible.

Nevertheless, at least for binary black hole mergers (BBHMs), these surveys will probe very large cosmic volumes, since they will be able to collect events from almost the full sky, with so high sensitivity that their horizon will reach very high redshifts ($z \sim 10$). Therefore, it will be possible to use them for high-precision measurements of cosmological parameters, either by considering their angular power spectrum (Namikawa, Nishizawa & Taruya 2016; Libanore et al. 2021; Namikawa 2021; Mukherjee et al. 2021a; Libanore et al. 2022; Yang & Hu 2022; Zhu et al. 2022), or cross-correlating their distribution with that of other tracers (Scelfo et al. 2018, 2020; Balaudo et al. 2022; Mukherjee et al. 2022; Libanore, Liguori & Raccanelli 2023; Scelfo et al. 2023). Note that GW surveys do not allow for a direct measurement of the redshift at which the merger took place: thus, a tomographic analysis of the sources distribution has either to assume the background cosmology (as it is done e.g. for LIGO–Virgo–KAGRA catalogues Abbott et al. 2016, 2019; Collaboration & Collaboration 2021; Collaboration & the Virgo Collaboration 2021), to use cross-correlations or other statistical techniques (e.g. in Del Pozzo 2012; Mukherjee et al. 2021a; Scelfo et al. 2022), or to rely only on luminosity distance (adopted, e.g. in Namikawa et al. 2016; Libanore et al. 2021) as a radial position indicator. All these approaches have been so far considered in the literature and have their specific advantages and drawbacks. Methodologies that infer the redshift and combine it with the luminosity distance of GW events also allow for interesting generalizations of the Alcock–Paczynski test (Mukherjee & Wandelt 2018). Moreover, it has been pointed out that GW spatial clustering could be used to detect the presence of primordial black hole (PBH) merger events in a catalogue (Raccanelli et al. 2016; Scelfo et al. 2018, 2020; Libanore et al. 2023).

★ E-mail: matteo.peron@unipr.it

© 2024 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

All these applications entail the study of GW sources as large-scale structure (LSS) tracers; therefore, a key ingredient for their implementation is a detailed understanding of the cosmological bias properties of the merger. The bias defines how sources cluster with respect to the underlying dark matter (DM) field; in the case of astrophysical merger events, be they resolved or unresolved, their bias $b_{\rm m}$ is linked with the underlying bias of both their host galaxies and host DM haloes. Based on how this link is modelled, a plethora of different prescriptions for $b_{\rm m}$ has been proposed in the literature. They range from assuming linear dependency on the host galaxy bias $b_{\rm g}$ (Oguri 2016), to redshift-dependent power laws (e.g. Vijaykumar et al. 2020; Cañas Herrera, Contigiani & Vardanyan 2021; Mukherjee et al. 2021a, b; Libanore et al. 2022), to more refined parameterizations, accounting for the properties of the binary and/or its host galaxy (e.g. Boco et al. 2019; Calore et al. 2020; Mukherjee & Silk 2020; Scelfo et al. 2020; Capurri et al. 2021; Libanore et al. 2021; Bellomo et al. 2022).

This significant variety in bias models and assumptions is an issue that can propagate to generate systematic and statistical uncertainties in cosmological studies of GW clustering. In this work, our main goal is to address and clarify this issue, focusing on a detailed study of the cosmological bias of stellar black hole merger events. In order to capture the complexity of the binary formation process as accurately as possible, we rely on state-of-the-art results, obtained by combining binary black hole catalogues from population synthesis models and the outputs of hydrodynamical, cosmological simulations. More specifically, our data set is based on the EAGLE cosmological, hydrodynamical simulation (Schaye et al. 2015) populated with stellar black hole binaries through the MOBSE population synthesis code (Mapelli et al. 2017; Giacobbo, Mapelli & Spera 2018) and evolved to get the final distribution of binary mergers, as described by Mapelli et al. (2017), Mapelli & Giacobbo (2018), and Artale et al. (2019, 2020) (hereafter, A19 and A20).

We use these data to build models and predictions for the bias of BBHMs, in different ways. First, we implement a suitable estimator of the merger power spectrum and directly measure the bias from the simulations; second, we use a halo occupation distribution (HOD) approach to build a semi-analytical model of the bias (following Libanore et al. 2022), based on the measured abundance of mergers in the simulations, as a function of the galaxy stellar mass and star formation rate. After testing the consistency of the results for internal validation purposes, we also compare them with the GW bias model of Scelfo et al. (2020) and we make an accurate assessment of how the starting assumptions in the different methodologies impact the final bias prediction. Finally, we use the input simulations to train a neural network (NN) to map the relation between the number of GW sources and parameters describing the properties of their galaxy or DM halo hosts. The main purpose in this case is to understand whether the general models used in semi-analytical approaches capture all the relevant Physics or could be improved by including extra parameters.

The paper is organized as follows. Section 2 contains a summary description of the simulations used for our analysis. In Section 3, we briefly review some key aspects of the bias model that we consider in this work. In Section 4, we describe and provide the main results of our bias estimator and discuss the formulation and calibration of semi-analytical models for GW bias and give details on the implementation. In Section 5, we characterize the architecture of our NN and we describe its possible applications. Conclusions can be found in Section 6.

2 MOCK DATA SET

Our study is performed on a simulated data set of BBHM events obtained populating two sets of the EAGLE simulations (Schaye et al. 2015) with merger events drawn from the MOBSE population synthesis code (Giacobbo et al. 2018), introduced first in A19 and A20.

We mainly use the EAGLE simulation with a box size of 100 Mpc and 1504^3 gas and DM particles, in which the cosmology is fixed to the *Planck 2013* results (Ade et al. 2014). We consider seven redshift snapshots, specifically at $z \in \{0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6\}$. For each, we use the following:

(i) the full DM particle snapshot (namely, the position of all the DM particles in the box) to estimate the matter density field in Section 4.1;
(ii) the host galaxy and host halo catalogues (describing the physical properties of the galaxies and haloes that host the mergers), which are a key ingredient for the semi-analytical model described in Section 4.2 and for the implementation of the machine-learning algorithm in Section 5.

Moreover, we consider a BBHM catalogue obtained by post-processing the previously described snapshots with MOBSE. While MOBSE can only provide stellar binary catalogues based on models that link galaxy properties to stellar populations, the post-processing procedure employed by A19 and A20 combines the generated binaries with the EAGLE stellar particles. This effectively populates each galaxy inside the hydrodynamical simulation with stellar binaries consistent with its properties, which then evolve into BBHM; we refer to the output catalogues of this procedure as EAGLE + MOBSE. We use the BBHM data set (which includes position and physical properties of all the merger events) to estimate the autopower spectrum and cross-power spectrum with DM in Section 4.1, and train and test our machine-learning algorithm in Section 5. Note that in the former case, we marginalize over all the properties of the mergers (e.g. progenitor masses, spins etc.), while for the latter we include them explicitly.

In the first part of our analysis, we develop an algorithm to estimate the power spectrum and clustering properties (i.e. the bias) of a mock data set. The EAGLE suite provides simulations with volumes of $(100\,{\rm Mpc})^3$, or smaller; such small volumes are due to the huge computational cost required to solve the full hydrodynamics, but it complicates the cosmological application, since the large linear scales are inaccessible from the simulation and non-linear effects become non-negligible, particularly at low redshift. As we discuss in Section 3, this makes it necessary to go beyond tree-level in the power spectrum computation and to jointly measure the linear bias b_1 together with higher order bias terms (see e.g. the review in Desjacques, Jeong & Schmidt 2018).

We validate the results of our algorithm by applying it to the study of DM halo clustering in the QUIJOTE simulation (Villaescusa-Navarro et al. 2020). QUIJOTE provides N-body DM-only snapshots at $z \in \{0,\ 0.5,\ 1,\ 2,\ 3\}$ inside a $L = 1000\,h^{-1}$ Mpc comoving box. Using QUIJOTE we have access to large, linear scales that allows us to directly extract b_1 and to compare the results of our b_1 measurement at non-linear scales with tree-level power spectrum and standard bias estimators to check the reliability of our pipeline. More detail on this validation test can be found in Appendix C.

The second part of our analysis relies on semi-analytical models. To calibrate them, we also use the EAGLE + MOBSE catalogues, in this case to extract the probability density functions (PDF) of galaxies

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

and merger events, as a function of the galaxy stellar mass and star formation rate.

Finally, we train and test an NN that reproduces the map between the number of mergers and properties of their host. For this, we use the aforementioned $L = 100\,\mathrm{Mpc}$ EAGLE + MOBSE data set and a smaller realization based on the 25 Mpc side with initial 752^3 gas and DM particles. In the following, we refer to these mock data sets as EAGLE100 and EAGLE25, respectively.

3 A SHORT OVERVIEW OF COSMOLOGICAL BIAS

Forthcoming GW detectors will access almost the full sky, but they will be characterized by a low sky-localization precision (see e.g. results by Iacovelli et al. 2022). Therefore, a power spectrum analysis of GW clustering will be able to probe only large-scale modes, which are in the linear regime. On such scales, for standard Gaussian initial conditions, the bias of a general DM tracer is well described by a single constant coefficient, b_1. In this work, we want to directly measure b_1 for GW sources from cosmological, hydrodynamical simulations to provide a benchmark for future studies.

The simulation we analyse, however, is necessarily limited to small cosmological volumes and contains scales ranging from mildly to fully non-linear: The size of the simulated box sets $k_{\rm min} = 2\pi 10^{-2}\,h\,\mathrm{Mpc}^{-1}$, while $k_{\rm max}$ is set by the Nyquist frequency, which, in turn, depends on how finely the mass distribution is sampled within the simulation. To avoid too large non-linearities, we only take into account scales up to $\sim 0.3\,h\,\mathrm{Mpc}^{-1}$, where perturbative bias models start to become significantly more complex. Here, several higher order parameters become non-negligible, thus we have to measure and marginalize over them if we want to estimate b_1.[1]

In this work, we adopt the bias convention detailed in Desjacques et al. (2018), which collects earlier results from, for example, McDonald & Roy (2009) and Chan, Scoccimarro & Sheth (2012). The expansion of a generic LSS tracer field $\delta_{\rm t}$ (e.g. DM haloes, $t = h$, galaxies, $t = g$, or BBHM, $t = m$), in terms of the matter density contrast δ up to third order, can be written as

$$\begin{aligned}\delta_t \sim\ & b_{1,\rm t}\delta + b_{\nabla^2\delta,\rm t}\nabla^2\delta + \epsilon\ + \\ &+ \frac{1}{2}b_{2,\rm t}\delta^2 + b_{K^2,\rm t}(K_{ij})^2 + \epsilon_\delta\delta\ + \\ &+ \frac{1}{6}b_{3,\rm t}\delta^3 + b_{\delta K^2,t}\delta(K_{ij})^2 + b_{K^3,t}(K_{ij})^3 + b_{\rm td,\it t}\,O^{(3)}_{\rm td}\ + \\ &+ \epsilon_{\delta^2}\delta^2 + \epsilon_{K^2}(K_{ij})^2\,, \end{aligned} \tag{1}$$

where, for convenience, we dropped the time τ and scale $\boldsymbol{k}$ dependencies of the matter density field δ and of the bias operators and parameters. The first three lines capture correspondingly increasing orders in the bias expansion and contain local ($b_{i=1,2,...}$, b_{K^2}), non-local ($b_{\nabla^2\delta}$, $b_{\rm td}$), and stochastic (ϵ_X) contributions. Local terms arise as a direct consequence of the equivalence principle, which states that in a free-falling frame of reference the leading gravitational observables for a comoving observer are given by the second-order derivatives of the metric tensor, which in the case of non-relativistic tracers reduces to the time–time component, the gravitational potential Φ. The density field δ is then given by the trace $\partial_i\partial_i\Phi$, while the tidal field K_{ij} is the trace-free part. Non-local terms arise when considering that a tracer does not form locally in space, but rather from the collapse of matter in a finite patch. Deviations from locality can be modelled into the bias expansion as convolutions of the local terms with kernels encompassing that patch, and these can be expanded to introduce higher derivative terms. Finally, stochastic terms arise from the small-scale perturbations, and are stochastic because they are not correlated over large distances, where the bias expansion is valid.

From equation (1), the tracer autopower spectrum, $P_{\rm tt}(k)$, and its cross-power spectrum with the matter field, $P_{\rm tDM}(k)$, can be directly derived up to next-to-leading order in the expansion. These include terms up to fourth order in the perturbative expansion of δ, and can be written as

$$\begin{aligned}P_{\rm tt}(k) = {}& b^2_{1,\rm t}\big(P_{\rm L} + 2P_{13} + P_{22}\big)(k)\ + \\ &+2b_{1,\rm t}\sum_{O\in\{\delta^2,\,K^2,\,\rm td\}} b_{\rm O,t}(F_{\rm O} + I_{\rm O})(k)\ + \\ &+\sum_{\rm O,\,O'\in\{\delta^2,\,K^2\}} b_{\rm O,t}b_{\rm O',t}I_{\rm OO'}(k)\ + \\ &+P_\epsilon - 2b_{1,\rm t}b_{\nabla^2\delta,\rm t}k^2P_{\rm L}(k)\,, \end{aligned} \tag{2}$$

$$\begin{aligned}P_{\rm tdM}(k) = {}& b_{1,\rm t}\big(P_{\rm L} + 2P_{13} + P_{22}\big)(k)\ + \\ &+\sum_{O\in\{\delta^2,\,K^2,\,\rm td\}} b_{\rm O,t}(F_{\rm O} + I_{\rm O})(k)\ - \\ &-b_{\nabla^2\delta,\rm t}k^2P_{\rm L}(k)\,, \end{aligned} \tag{3}$$

where $P_{\rm L}$ is the tree-level matter power spectrum, while P_{13} and P_{22} is the one-loop corrections due to non-linear evolution in perturbation theory (see e.g. Peebles 1980; Bernardeau et al. 2002, for a review). With respect to equation (1), we here introduced $b_{\delta^2} = b_2/2$. All the other terms define the tree-level and one-loop contributions to the matter power spectrum; we explicitly define them in Appendix A (for a complete derivation see Desjacques et al. 2018, and references therein).

We refer to the model in equations (2) and (3) as EFTofLSS and we use it as starting point for the mock-data bias analysis in Section 4.1. We also consider a simpler model, in which third-order and scale-dependent contributions are neglected. This is defined as

$$\begin{aligned}P_{\rm tt}(k) = {}& b^2_{1,\rm t}\big(P_{\rm L} + 2P_{13} + P_{22}\big)(k)\ + \\ &+2b_{1,\rm t}\sum_{O\in\{\delta^2,\,K^2\}} b_{\rm O,t}(F_{\rm O} + I_{\rm O})(k)\ + \\ &+\sum_{\rm O,\,O'\in\{\delta^2,\,K^2\}} b_{\rm O,t}b_{\rm O',t}I_{\rm OO'}(k) + P_\epsilon\,, \end{aligned} \tag{4}$$

$$\begin{aligned}P_{\rm tdM}(k) = {}& b_{1,\rm t}\big(P_{\rm L} + 2P_{13} + P_{22}\big)(k)\ + \\ &+\sum_{O\in\{\delta^2,\,K^2\}} b_{\rm O,t}(F_{\rm O} + I_{\rm O})(k)\,. \end{aligned} \tag{5}$$

We label this model as LIMD (Local In Matter Density), as it accounts only for local contributions of the matter field to the power spectrum.

In both equations (2) and (4), the P_ϵ term represents the first-order stochastic contribution. This is scale-independent and in first approximation corresponds to the shot noise $\bar{n}_t^{-1}$, where $\bar{n}_t$ represents the tracer density (Dekel & Lahav 1999). To reduce the dimensionality of the problem, we neglect higher order, subdominat,

[1] A more refined way to estimate b_1 from mock data, achieving higher precision and without including higher order coefficients, would be to rely on a separate Universe approach (Mead & Peacock 2014; Wagner et al. 2015; Li, Hu & Takada 2016) to extrapolate the currently available EAGLE + MOBSE results on larger scales. Unfortunately, the separate Universe methodology requires many simulations, generated for different input values of $\Omega_{\rm m}$, which are not available for GW sources. In this work, we indeed rely on a single realization, at fixed fiducial cosmological parameters (Ade et al. 2014).

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

stochastic terms by setting $\epsilon_\delta = \epsilon_{\delta^2} = \epsilon_{K^2} = 0$ and $P_\epsilon = \bar{n}_t^{-1}$. We also rewrite $b_{K^2,\mathrm{t}} = -2(b_{1,\mathrm{t}} - 1)/7$ to further reduce the number of free parameters in the analysis (Baldauf et al. 2012). The interested reader can find further details on this part of the analysis in Appendix B.

Higher order correlators (e.g. the bispectrum, Matarrese, Verde & Heavens 1997) are useful to measure $b_{1,t}$ on large scales (as it is done e.g. in Jung et al. 2022), and to break degeneracies between bias parameters at different orders. In the range of scales that we can safely access from the EAGLE simulation ($2\pi 10^{-2} \lesssim k \lesssim 0.3\,h\,\mathrm{Mpc}^{-1}$), they should be expanded beyond the tree-level as well. This would make their implementation particularly challenging; moreover, their application to real data analysis would be limited by the lack of sky-localization in GW detection. For these reason, we leave further investigation on this point for a follow-up work.

4 SIMULATION-BASED BIAS MODELLING

The linear bias parameter $b_{1,\mathrm{t=m}}$ of BBHMs can be estimated from the data sets presented in Section 2 by determining the tracer power spectra from the catalogues and fitting them with the models discussed in Section 3. We present our $b_{1,\mathrm{m}}$ estimates in Section 4.1 and we test our results against semi-analytical models in Section 4.2, as a counter-check of the reliability of this approach.

In this way, we are also able to understand which physical parameters are crucial to determine the clustering of the sources. Such analysis can then be compared with the outputs of the machine-learning algorithm described in Section 5, in which the relative importance of the parameters is tested inside the simulation itself.

4.1 Direct estimation on mock data

To measure the bias of any given tracer t in our mock data sets, we first extract the tracer autopower spectrum $P_{\mathrm{tt}}(k)$ and DM-tracer cross-power spectrum $P_{\mathrm{tDM}}(k)$, by relying on a straightforward estimator that can be summarized as follows:

(i) We estimate the number density field $\tilde{n}_x(\boldsymbol{r})$ (x can be either DM or the tracer t) by weighing the particle positions with a cloud-in-cell (CIC) scheme.

(ii) We compute the density contrasts $\tilde{\delta}_x(\boldsymbol{r}) = \tilde{n}_x(\boldsymbol{r})/\bar{\tilde{n}}_x - 1$ (the bar denotes spatial average) and we perform a fast Fourier transform (FFT) to get $\tilde{\delta}_x(\boldsymbol{k})$.

(iii) We deconvolve the CIC filter function from the estimated $\tilde{\delta}(\boldsymbol{k})$ to recover the actual Fourier transform of the field, $\delta(\boldsymbol{k})$.

(iv) We measure the spectrum estimator, $\tilde{P}_{xy}$, by averaging $\delta_x(\boldsymbol{k})\delta_y^*(\boldsymbol{k})$, inside bins in Fourier space and estimate the 1σ error bars on $\tilde{P}_{xy}(k)$ by taking the square root of the Gaussian variance of our results (Scoccimarro, Zaldarriaga & Hui 1999).

As a counter-check, we also implement a second power spectrum estimator, based on the Feldman–Kaiser–Peacock (FKP) prescription (Feldman, Kaiser & Peacock 1994). Further details on the technical implementation of the two estimators are discussed in A ppendix B.

Having obtained the estimates $\tilde{P}_{\mathrm{tt}}(k)$ and $\tilde{P}_{t\mathrm{dM}}(k)$, we then fit the models of equations (2)–(5) to our results via Markov chain Monte Carlo (MCMC). We assume a Gaussian likelihood for $\tilde{P}(k)$ and consider the parameter vectors $\boldsymbol{\theta}_{\mathrm{EFTofLSS}} = (b_{1,\mathrm{t}},\, b_{2,t},\, b_{\mathrm{td},t},\, b_{\nabla^2\delta,\mathrm{t}})$ and $\boldsymbol{\theta}_{\mathrm{LIMD}} = (b_{1,\mathrm{t}},\, b_{2,t})$, for the EFTofLSS and LIMD models, respectively; we then take a uniform prior in the interval $b_{1,\mathrm{t}} \in [0,\ 10]$ for the linear bias parameter and uniform improper priors on all other parameters. The choice of $b_{1,\mathrm{t}} > 0$ is physically motivated, as negative values would mean that tracers cluster in regions where DM is less dense. The same argument cannot be applied as straightforwardly to the other bias parameters, as their effect on clustering is not as trivial: for this reason we consider uninformative priors so that the chains can freely explore the parameter space without skipping possible high-probability density regions. We combine the autopower spectra corrected for the shot noise (that is, we subtract $P_\epsilon = \bar{n}_t^{-1}$) with the cross-power spectra and truncate the analysis at a given scale k_{NL} (specified later), chosen to exclude the highly non-linear regime where perturbation theory fails. MCMC sampling of the posterior is performed with the `emcee`[2] package (Foreman-Mackey et al. 2013), while for the computation of the bias operators I_{O}, F_{O}, and $I_{\mathrm{OO'}}$ that appear in equations (2)–(5) we use `velocileptors`[3] (Chen, Vlah & White 2020; Chen et al. 2021), a package designed to evaluate perturbative models of the power spectrum of biased tracers, like the ones we consider. To constrain $b_{1,\mathrm{t}}$, we treat all the non-linear bias parameters as nuisance parameters and marginalize over them. Our final results are maximum a posteriori estimates and we use the 18th and 84th percentiles to define our $b_{1,\mathrm{t}}$ credible interval (see Figs C4 and C5). The reliability of our bias estimator is first tested on DM halo catalogues from the QUIJOTE simulations, for which we can easily compare our results with theoretical expectations (see Appendix C).

Table 1. $b_{1,\mathrm{m}}$ estimates obtained via MCMC on EAGLE + MOBSE data sets.

z	EFTofLSS $b_{1,\mathrm{m}}$	LIMD $b_{1,\mathrm{m}}$
0	$1.00^{+1.92}_{-0.44}$	$1.29^{+0.09}_{-0.84}$
1	$1.21^{+1.78}_{-0.54}$	$1.81^{+0.07}_{-0.93}$
2	$1.35^{+2.36}_{-0.56}$	$2.23^{+0.13}_{-0.88}$
3	$1.73^{+2.80}_{-0.65}$	$2.67^{+0.22}_{-0.91}$
4	$2.35^{+3.11}_{-0.65}$	$3.11^{+0.24}_{-0.96}$
5	$4.50^{+1.76}_{-2.14}$	$3.49^{+0.33}_{-0.99}$
6	$5.73^{+1.49}_{-2.81}$	$3.80^{+0.56}_{-0.89}$

We then apply our algorithm to estimate the linear bias of BBHM at redshifts $z \in \{0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6\}$. In this analysis, we truncate the power spectra at $k_{\mathrm{NL}} = 0.25\,h\,\mathrm{Mpc}^{-1}$ for $z = 0$, and $k_{\mathrm{NL}} = 0.35\,h\,\mathrm{Mpc}^{-1}$ for all other redshifts. The results of the analysis are summarized in Table 1 and Fig. 1, which, respectively, show and plot the values of $b_{1,\mathrm{m}}$ and the corresponding errors in the chosen redshift bins.

For practical purposes and applications in follow-up analysis, we provide the following power-law fit to the bias estimated from simulations:

$$b_{1,\mathrm{m}} = 1.20\,(1+z)^{0.59}\,. \tag{6}$$

The parameters in the previous equation are calibrated on the LIMD results, since we find that the LIMD model is favoured by the Bayes Information Criterion over the EFTofLSS one. This makes intuitive sense, the number of Fourier modes accessible at the perturbative regime of the EAGLE100 is of $\mathcal{O}(1)$: with so little data, complex models such as EFTofLSS that have several nuisance parameter will give a posterior with high variance after marginalization, while the LIMD, despite being inaccurate for not accounting for non-local effects, can still yield a posterior in the high-

[2] https://github.com/dfm/emcee.
[3] https://github.com/sfschen/velocileptors.

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

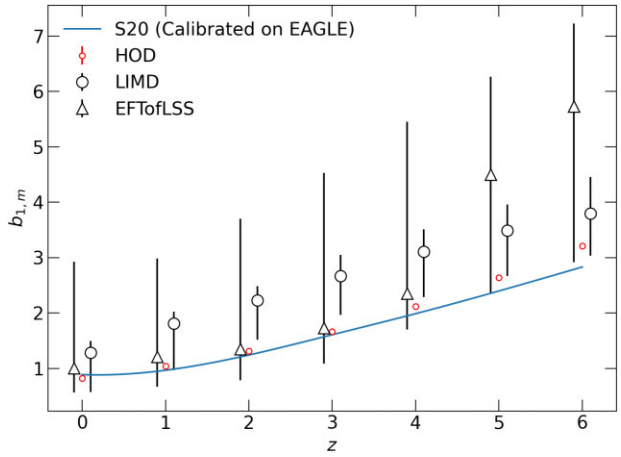

Figure 1. Bias estimates computed with the different methods described the text: MCMC results from Section 4.1 and Table 1 in the LIMD (black circles) and EFTofLSS (black triangles) cases, semianalitycal HOD model from Section 4.2 (red circles) and updated merger rate-weighted results from Section 4.3 (S20, blue line).

probability region with just enough variance to fit the true underlying model.

4.2 Semi-analytical model

To validate the results of the previous section, we also evaluate the bias with a different approach, which exploits information extracted from the simulations as an input to build an HOD model (see e.g. Peacock & Smith 2000; Berlind & Weinberg 2002; Karagiannis et al. 2018; Libanore et al. 2021) and connect the BBHM bias with the bias of their host galaxies and DM haloes

In Libanore et al. (2021), a similar technique was firstly applied to the study of merger bias. With respect to the usual HOD approach (Peacock & Smith 2000; Berlind & Weinberg 2002) – aimed at modelling the bias of galaxies from the distribution of their host DM haloes – this now requires an extra step to account for the BBHM distribution as a function of the host galaxy features; all the relevant distributions are calibrated using our mock EAGLE + MOBSE data set, also analysed in the previous section. Compared to Libanore et al. (2021), we now adopt an updated simulation setup, using EAGLE100 (while Libanore et al. 2021 used data from EAGLE25). We also use the full mock data set (the same as in Section 4.1), while in Libanore et al. (2021) specific selection effects due to the ET signal-to-noise ratio were implemented.

We now discuss the aforementioned procedure more in detail. First of all, we need to compute the bias of the host galaxies, which requires the following three ingredients:

(i) the halo mass function $\mathrm{d}n(z)/\mathrm{d}M_\mathrm{h}$ in each snapshot, i.e. the number of haloes per mass bin. We consider six halo mass bins with M_h uniformly spaced in log-space between 10^9 and $4 \times 10^{14}\,\mathrm{M}_\odot$.

(ii) the linear halo bias $b_{1,\mathrm{h}}(M_\mathrm{h}, z)$, for which we rely on the parametrization derived in Tinker et al. (2010) using the peak-background split (PBS) approach. In Appendix C – Figs C3 and C2 – we compare this parametrization with the actual estimates of $b_{1,\mathrm{h}}$ from simulations at $z = \{0, 3\}$, obtained from the algorithm described in the previous section.

(iii) the galaxy HOD $\langle N_\mathrm{g}(M_*, \psi, z)|M_\mathrm{h}\rangle$, namely the probability of finding N_g galaxies having stellar mass M_* and star formation rate ψ inside a halo of mass M_h in the z snapshot. The number per bin is directly estimated from the simulations; we consider 15 stellar mass bins between $M_* = 10^7$ and $10^{12}\,\mathrm{M}_\odot$ and 15 star formation rate bins between $\psi = 10^{-5.5}$ and $10^{2.5}\,\mathrm{M}_\odot\mathrm{yr}^{-1}$.

We note that in Libanore et al. (2021) a similar procedure was applied by fitting data with the Tinker et al. (2010) halo mass function and the EAGLE HOD analytical prescription (Artale et al. 2018). We verified that such prescription holds as a first-order approximation, but using the halo and galaxy PDFs directly measured on simulations is more reliable when considering bins with a small number of data.

By combining all these ingredients, we compute the galaxy linear bias per stellar mass and star formation rate bin at each redshift as

$$b_{1,\mathrm{g}}(M_*, \psi, z) = \frac{\int \mathrm{d}M_\mathrm{h}\, b_{1,\mathrm{h}}(M_\mathrm{h}, z)\frac{dn(z)}{\mathrm{d}M_\mathrm{h}}\langle N_\mathrm{g}(M_*, \psi, z)|M_\mathrm{h}\rangle}{\int \mathrm{d}M_\mathrm{h}\,\frac{dn(z)}{\mathrm{d}M_\mathrm{h}}\langle N_\mathrm{g}(M_*, \psi, z)|M_\mathrm{h}\rangle}, \quad (7)$$

where the numerical integration is performed between the values of M_h^{min} and M_h^{max} that define the halo mass bins.

As mentioned, the semianalytical model we build to describe the merger bias adds an extra layer to the analysis discussed up to this point, namely the galaxy occupation distribution $\langle N_\mathrm{m}(z)|M_*, \psi\rangle$, which we define as the probability of finding N_m mergers in a galaxy as a function of its stellar mass M_* and star formation rate ψ. We estimate $\langle N_\mathrm{m}(z)|M_*, \psi\rangle$ from the EAGLE + MOBSE simulation and finally model the linear bias of the merger as

$$b_{1,\mathrm{m}}(z) = \frac{\int \mathrm{d}M_* \int \mathrm{d}\psi\, b_{1,\mathrm{g}}(M_*, \psi, z)\frac{\mathrm{d}N_\mathrm{g}(z)}{\mathrm{d}M_*\mathrm{d}\psi}\langle N_\mathrm{m}(z)|M_*, \psi\rangle}{\int \mathrm{d}M_* \int \mathrm{d}\psi\,\frac{dN_\mathrm{g}(z)}{\mathrm{d}M_*\mathrm{d}\psi}\langle N_\mathrm{m}(z)|M_*, \psi\rangle}, \quad (8)$$

and also in this case the numerical integration is performed between the M_* and ψ bins boundaries. The galaxy number density $\mathrm{d}n_\mathrm{g}(z)/\mathrm{d}M_*d\psi$ is computed by dividing the number of galaxies in each bin $N_\mathrm{g}(M_*, \psi, z)$ by the width of the $\Delta M_* \Delta\psi$ bin and the volume of the simulation box, L^3. We checked that marginalizing over the star formation rate dependence before computing the galaxy and merger bias (i.e. dividing the galaxies and merger only in dependence of M_h, M_*) has a small impact on the final results. Indeed, as we discuss in detail in Section 5, the merger distribution properties mainly depend on the stellar mass (this is in agreement to what was initially pointed out in A19 and then confirmed in Santoliquido et al. 2022). The HOD contains an intrinsic uncertainty due to the theoretical assumptions in the model (e.g. assuming dependence only on the halo mass, stellar mass and star formation rate). With this caveat in mind, we verified that the statistical error generated by propagating[4] the uncertainties on $\langle N_\mathrm{g}(M_*, \psi, z)|M_\mathrm{h}\rangle$ and $\langle N_\mathrm{m}(z)|M_*, \psi\rangle$ in equations (7) and (8) is negligible compared to the error bars provided by the numerical estimate of $b_{1,m}$ in the previous section. The internal comparison between the semi-analytical linear bias of the mergers estimated from equation (8) and the results from Section 4.1 is displayed in Fig. 1. Results are in good agreement between these two methods and with the close-to-linear trend previously obtained by Libanore et al. (2021, 2022).

[4]The error estimate was computed by assuming that N_g and N_m are independent, Gaussian random variables. We apply the standard error propagation formulae separately on the numerator and denominator, assuming that the error contribution from $\mathrm{d}n/d M_\mathrm{h}$ and $b_{1,\mathrm{h}}(M_\mathrm{h}, z)$ is negligible. The ratio between these two quantities is described by a Cauchy distribution, for which the expected value and variance are undefined; we overcome this problem relying on the prescriptions in Marsaglia (2006).

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

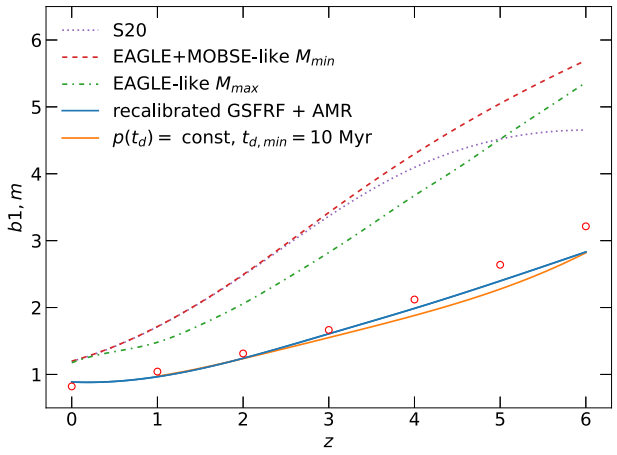

Figure 2. HOD bias values from equation (8) (red circles) compared with the merger rate-weighted biases computed via equation (10) under different prescriptions. The dotted line uses the same setup as S20, while we changed the lower and upper halo mass integration limits in the dashed and dot–dashed lines, respectively. We obtained the blue, thick line by re-calibrating the star formation rate and abundance matching results on the EAGLE catalogues and expressing the time delay distribution as $p(t_d) \propto t_d^{-1}$. The orange, thick line relies instead on the same EAGLE-like configuration, but using a constant time delay distribution, above $t_{d,\,min} = 10$ Myr. More detail in the text.

4.3 Comparison with the merger-rate-weighted bias

The main goal of our analysis is to understand how much the BBHM bias is affected by the properties of the hosts. This is a key information to understand whether the clustering can be used to disentangle different host populations (and possibly formation channels), as well as to check how much our results depend on the specific astrophysical models assumed in the simulations. With this in mind and to verify the robustness of our results, we re-estimated the bias following a completely independent approach with respect to the one adopted in previous sections. This alternative methodology is based on the analysis depicted in Boco et al. (2019, hereafter B19) and Scelfo et al. (2020, hereafter S20) where the BBHM bias is computed by weighting the bias of the hosts through the merger rate, which, in turn, relates with the host properties, modelled from theoretical assumptions and calibrated on real data (a similar approach is also followed by Bellomo et al. 2022).

The parametrization used in B19 and S20 is different with respect to the assumptions made by the authors of A19 to simulate the BBHM catalogues we rely on. This propagates to the bias computation, leading to discrepancies between the results of the two procedures (particularly at high redshift). Fig. 2 illustrates so, showing that the bias in S20 (dotted line), as well as the bias in Scelfo et al. (2018), is at face value consistent neither with each other nor with our results, represented by the red circles. To shed light on this possible issue, we focus on S20 and investigate the reason for the discrepancy. We model the bias according to the procedure described in S20, to which we defer the reader for details, while we limit ourselves here to briefly recap the necessary building blocks. In the same spirit as equations (7) and (8), the bias is calculated convolving the halo mass function and halo bias with an appropriate kernel, which describes the number distribution of BBHM, conditioned over a few properties of the BBHM and of the haloes and galaxies which host them: the star formation rate ψ, the metallicity Z, the halo mass and galaxy mass. The last two can be expressed in terms of ψ, via the abundance-matching relations (AMR) of Aversa et al. (2015). B19 and S20 motivate in detail that the kernel can be characterized via the BBHM merger rate $\mathcal{R}_{\rm BH}(z, \log_{10}\psi, \mathcal{M}_{\rm c})$, where $\mathcal{M}_{\rm c}$ is the BBHM chirp mass. This is defined in S20 as

$$\mathcal{R}_{\rm BH}(z, \log_{10}\psi, \mathcal{M}_{\rm c}) \propto \int {\rm d}t_{\rm d}\, p(t_{\rm d})\, \psi\, \frac{{\rm d}^2 N(\log_{10}\psi, t - t_{\rm d})}{{\rm d}}\, V \times \\ \times \int {\rm d}\log_{10} Z\; p(\log_{10} Z | t - t_{\rm d}, \psi)\, p(\mathcal{M}_{\rm c}|Z). \tag{9}$$

In the previous expression, the proportionality is related with the observed cosmic volume V and with the fraction of BBHMs capable to merge within an Hubble time. The computation depends on the number N of BBHMs that takes place inside V, depending on the star formation rate ψ in the host galaxy, and on the time delay $t_{\rm d}$ between the binary formation and merger; all the distributions are calibrated on data in S20. We account for the probability distribution of the chirp mass $\mathcal{M}_{\rm c}$, of the host galaxy metallicity Z and of the time dely $t_{\rm d}$. We model the BBHM chirp mass distribution as a function of the mass of the progenitor star and the metallicity of the host galaxy, as it is done in B19. We adopt the initial mass function from Chabrier (2003) and we relate it to the black hole mass that the star forms as in Spera & Mapelli (2017). The galaxy distribution is modelled according to the galaxy star formation rate function parametrized as a Schechter function in Mancuso et al. (2016). All the quantities in the analysis are estimated at the binary formation time: This requires us to introduce a reasonable distribution for the time delay $t_{\rm d}$ between this moment and the time of the merger. The isolated binary evolution predicts the time delay distribution $p(t_{\rm d}) \propto t_{\rm d}^{-1}$, with a minimum time delay cutoff of $t_{d,\rm min} = 50$ Myr, which is used in B19 and S20. In general, the exact form depends on several physical processes such as mass transfer, kicks, stellar winds, and orbital separation. On the other hand, the delay-time distribution changes depending on galaxy properties. Early-type galaxies present a flatter time delay distribution for BBHM compared with late-type galaxies (see A19, fig. 10, bottom panel).

Therefore, in our analysis, we test two independent cases: $p(t_{\rm d}) \propto t_{\rm d}^{-1}$ (to be consistent with S20), and a uniform distribution $p(t_{\rm d}) \equiv$ const that recalls the one implemented in the simulations by A19 analysed in Section 4.1. The specific choice of $p(t_{\rm d})$ may largely affect the final results on the black hole merger rate (Santoliquido et al. 2021), but how this propagates to the clustering of events, has not been discussed yet. We anticipate that our analysis shows that different choices of $p(t_{\rm d})$ have a limited impact on the bias.

Piecing everything together we estimate the merger bias through

$$b_{1,\rm m}(z) = \frac{\int {\rm d}\log_{10}\psi\; b_{1,\,\rm g}(z,\psi) \int {\rm d}\mathcal{M}_{\rm c}\; \psi\; \mathcal{R}_{\rm BH}(z, \log_{10}\psi, \mathcal{M}_{\rm c})}{\int {\rm d}\log_{10}\psi \int {\rm d}\mathcal{M}_{\rm c}\; \psi\; \mathcal{R}_{\rm BH}(z, \log_{10}\psi, \mathcal{M}_{\rm c})}\,, \tag{10}$$

where the galaxy bias as a function of ψ is computed as in Aversa et al. (2015), i.e. $b_{1,\,g}(\psi, z) = b_{1,\,\rm h}(M_{\rm h}, z)$, where the halo bias is the same we discussed in the previous section.

Different prescriptions on the models and parameters in equation (10) can lead to very different bias results. Tuning these quantities to different values, we obtain all the other lines shown in Fig. 2. The figure shows that, by matching the halo masses and star formation distributions to the ones estimated from EAGLE, we can force the results of equation (10) to be consistent with the bias estimates in Section 4.2.

The reasons underlying the difference between our bias estimate in equation (8) and S20 bias estimate in equation (10) can be summarized as follows:

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

(i) The results in S20 include detector selection effects via a signal-to-noise ratio distribution; to relate them to our analysis (which involves the intrinsic signal of the full merger population), we need to drop this term.

(ii) In S20, the lower bound of the star formation rate integral is directly related to a lower bound on the mass of the host galaxy via the AMR. At every redshift, this lower bound $M_{\rm min}$ is chosen such that the integral converges. Instead, in EAGLE + MOBSE at every redshift only galaxies with stellar mass $\gtrsim 10^7\,{\rm M}_\odot$ are included because of mass resolution limits (see e.g. the discussion in Mapelli et al. 2018). If the same cut is imposed in S20 (dashed line in Fig. 2) the characteristic flattening S20 had at high-redshift partially disappears, since the least biased objects are removed from the distribution.

(iii) The opposite effect, i.e. a decrease in the bias due to the absence of the most biased objects, takes place when we remove host galaxies with mass $\gtrsim 10^{12}\,{\rm M}_\odot$. This is not due to a physically motivated effect, but to the fact that the EAGLE simulation is quite limited in size, and does not contain larger objects. To circumvent this limitation, larger cosmological, hydrodynamical simulations or self-consistent extrapolation to higher massess are required.

(iv) Finally, we have to take into account that both the galaxy star formation rate function of Mancuso et al. (2016) and the AMRs provided in Aversa et al. (2015) do not fully agree with the corresponding quantities extracted from EAGLE. Re-calibrating these functions on the EAGLE catalogues, the bias estimated from equation (10) turns out to be in broad agreement with our HOD results from equation (8) (blue thick line in Fig. 2).

As anticipated, both to test how strong the dependence of the bias on the time delay distribution is, and to check consistency with the EAGLE + MOBSE simulation, we repeat the calculation using a different $p(t_{\rm d})$. We pick a flat distribution with a minimum time delay of $t_{d,\rm min} = 10\,{\rm Myr}$ from the distributions considered in Santoliquido et al. (2021); the two bias distributions ($p(t_{\rm d}) \propto t_{\rm d}$, and $p(t_{\rm d}) \equiv$ const) bracket the one used in EAGLE + MOBSE. In this way, we obtain the orange, thick line in Fig. 2. It is evident from the figure that the choice of time delay distribution has a negligible impact in the determination of the amplitude and redshift evolution of the BBHM bias; therefore, as we anticipated, we conclude that the BBHM clustering is only marginally affected by the specific choice of time delay distribution.

5 PAINTING EMULATOR AND FEATURE IMPORTANCE

5.1 Machine-learning methodology

The numerical simulations allow us to verify which physical parameters are the most relevant in the modelling of BBHM distribution and their spatial clustering. We train an NN to assign to each galaxy a number of BBHM events, based on the galaxy and the host halo properties. By varying the selection of input features, a set of NN can then be used to study their relative importance, and, when the informative parameteres have been identified, the NN can be used to quickly paint them on N-body or cosmological, hydrodynamical simulations. We notice how the BBHM distribution, not their spatial clustering, was previously studied with different methods in A19. Our results are in agreement with theirs.

We start by noticing that the BBHM number has a strong dependence on the galaxy mass $M_\star$.[5] To account for it, we fit the pairs ($M_\star$, $N_{\rm BBHM}$) of the training set with a cubic polynomial and calculate the fractional residuals dividing the BBHM number data by the fitting function, in order to remove the main scaling. This is not only motivated by speeding up the NN training, but also from the fact that the prediction from the simple cubic polynomial regression already leads to a reasonable approximation of the power spectrum. We trained different sets of NNs, as described below, on either the raw BBHMs number, and the residuals we just discussed, finding by direct inspection that the latter produces better results.

A first approach, established in Kamdar, Turk & Brunner (2016), finds an univocal link between the input variables and the number of BBHMs, i.e. a single combination of stellar mass, star formation rate, metallicity, etc., is always associated to a single BBHM number. It is known (Kamdar et al. 2016; Agarwal, Davé & Bassett 2018; De Santi et al. 2022) how this procedure tends to successfully reproduce the average properties of the target parameter, but fails to reproduce its distribution. In our case, this is driven by the fact that the BBHM have an intrinsic scatter due to both the probabilistic nature of the events, and the baryonic physics that is not easily captured by galaxy properties (Stiskalek et al. 2022). This means that two different galaxies that have the same average properties may contain a different number of BBHMs. The second approach circumvents this issue by training an NN to assign to each galaxy the BBHM PDF, using a suitable parametrization (Nix & Weigend 1994; Villaescusa-Navarro et al. 2021a; Stiskalek et al. 2022). Each galaxy is then populated with BBHMs drawing a single realization from the associated PDF. We will hereafter dub these two approaches *deterministic* and *statistical*.

Summarizing, we train two deterministic networks on either the BBHM number or the BBHM residuals, three statistical networks, each parametrizing a normal, lognormal or Poisson distribution, on the BBHM number, and two other statistical network, with normal or lognormal distribution, on the BBHM residuals. We do not apply the Poisson distribution to the residuals since this is not meaningful for PDF obtained as the ratio of two number of events.

Both the deterministic NN and the statistical NN share the same architecture for the input and hidden layers, while the output layers differs. In the former case, we have a single node output with `elu` + 1 activation both when we are fitting for the BBHM number and the fractional residuals. In the case of normal and lognormal distributions, we use a two node output: The first is interpreted as mean and the second as standard deviation of the variable (or logarithm of the variable, respectively). In the case of Poisson distribution, the single output is the mean of the distribution. In all output nodes, an `elu` + 1 activation function is used again, with the excepton of the mean of the lognormal distribution which has a linear activation. For the deterministic network, we use the mean square errors as loss function, while in the statistical network the likelihood of the parametrized distribution is used instead. As optimizer, we employ `Adam` (Kingma & Ba 2014) with a cyclical learning rate (Smith 2015), and we apply a dropout (Srivastava et al. 2014) to all hidden layers. We tune the number and hidden layers, nodes, dropout rate and activation functions, setting for 3 hidden layers of 128 nodes, a dropout rate of 0.1 and `selu` (Klambauer et al. 2017) activation functions.

We randomly divide the catalogues from EAGLE100 into training (60 per cent), validation (20 per cent), and test (20 per cent) sets. This leads to a spatially inconsistent distribution of the galaxies of each set; while we can assume that each set is a fair sample of the BBHM PDF, none of them can be used by themselves to calculate a meaningful power spectrum. Thus, the EAGLE25 is used to test the power spectrum reconstruction instead.

[5] We define the galaxy mass as the sum of the bounded stellar particle masses. The same argument applies to any other definition of galaxy mass.

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

Table 2. Parameters in the EAGLE catalogue used as input features for the NN (McAlpine et al. 2016). If the parameter is used in the text, the string used in the EAGLE database is paired with the symbol we use. The merger tree info are the value of said parameters for the object progenitors (if present) along the main branch within the 16 snapshots preceding the merger. Note that a 'group' in the simulation matches a DM halo, while a 'subgroup' is associated with a galaxy.

	Merger tree info	Description
Host galaxy parameter		
MassType_Star, $M_\star$ [$M_\odot$]	✓	Total stellar mass
SF_Metallicity, Z	✓	Metal mass fraction in the star-forming gas
StarFormationRate, ψ [$M_\odot$ yr^{-1}]	✓	Total star formation rate
Vmax, V_{max} [km s^{-1}]	✓	Maximum of the circular velocity rotation curve
StellarVelDisp [km s^{-1}]	✗	Average one dimensional star velocity dispersion
TotalEnergy [$M_\odot$ km^2 s^{-2}]	✗	Total energy of the galaxy
KineticEnergy	✗	Total kinetic energy
Host halo parameter		
GroupMass, M_H [$M_\odot$]	✓	Total firends-of-friends mass
Group_M_Crit200 [$M_\odot$]	✓	Total mass within the radius containing a density 200 times the critical density of the Universe
Group_M_Crit500 [$M_\odot$]	✓	Total mass within the radius containing a density 500 times the critical density of the Universe
Group_M_Crit2500 [$M_\odot$]	✓	Total mass within the radius containing a density 2500 times the critical density of the Universe

Since our goal is to verify how all the information on the BBHM distribution is encoded in a few parameters, we use a rather agnostic approach in choosing our baseline feature set and include the wide range of halo and galaxy parameters shown in Table 2, out of the scalar quantities provided in the EAGLE simulation.

5.2 Validation of the NN and its architecture

We start by discussing the necessity to move past the description of the average quantity in the context of ML-based painting of simulations, as also pointed out in Stiskalek et al. (2022). As anticipated, a limited amount of physical properties of the host halo and, in our case, of the host galaxy, fail to describe *the* single realization of the painted label. Instead, they characterize the label's PDF. Employing the NN to learn the PDF, rather than the single realization of the label, not only allows us to better reconstruct its variance, which is often underestimated (Kamdar et al. 2016; Agarwal et al. 2018; De Santi et al. 2022), but also to better recover properties related to the *average* label distribution.

First, we can test the NN ability to reproduce a BBHM distribution that is statistically equivalent to the test-set one, dubbed *true* in the following. In the two panels of Fig. 3, we compare how well our two architectures reproduce the true distribution as a function of the host galaxy mass. The host galaxy mass is singled out among all features because it correlates the most with the target BBHM distribution (see the discussion in Section 5.3).

We divide the events in 50 logarithmically spaced galaxy mass bins. In each, the average number of events is reproduced successfully both using the deterministic and the statistical network. However, in most bins, the former generates a distribution which underestimates the true scatter in the bin. In practice, galaxies that should host an extreme (either high or low) number of events are instead more likely populated with a number of BBHM closer to the average when the deterministic NN is used. The statistical network, instead, increases the variance of the proposal distribution from which the predictions are drawn when the population of events falls in a broad spectrum. This means that, on a galaxy-by-galaxy basis, the prediction of this NN might differ substantially from the true one, but in such a way that the whole distribution turns out to be more closely followed. The same behaviour appears in the global distribution of BBHM per galaxy shown in Fig. 4. To provide a quantitative metric of this intuition, we use the *k*-sample Anderson–Darling test

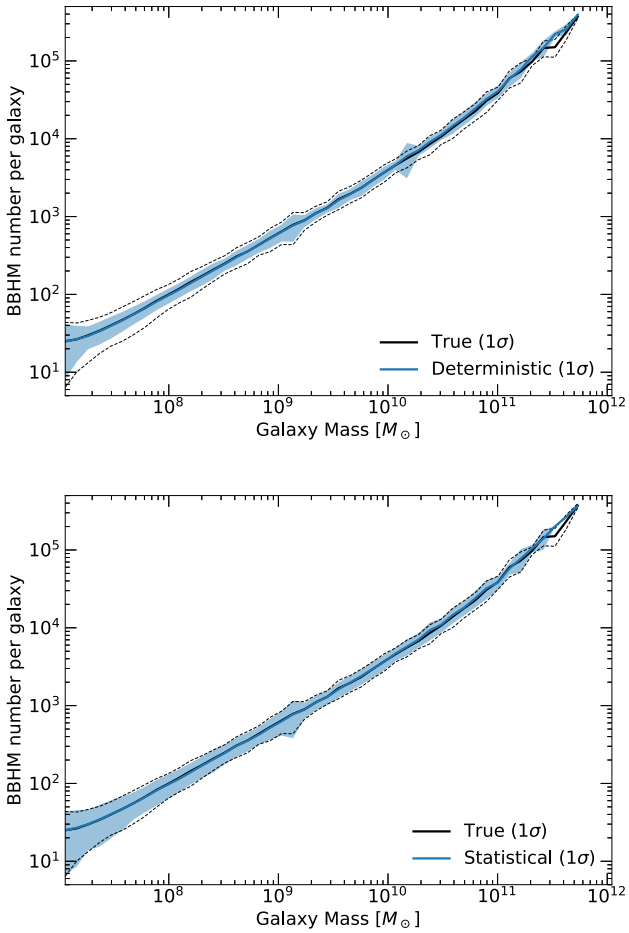

Figure 3. BBHM number distribution as a function of the host galaxy mass. The black line describes the 'true' distribution from the test set, while the blue line describes the NN result. The top panel refers to the deterministic architecture, while the bottom panel relates with the PDF (statistical) one.

(Scholz & Stephens 1987), whose score (ADS) quantifies how close the distributions underlying two samples are to each other. In other words, the ADS is related to the likelihood that the two samples are drawn from the same distribution. In the following, we will always calculate the ADS score of a sample against the true one. While the

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

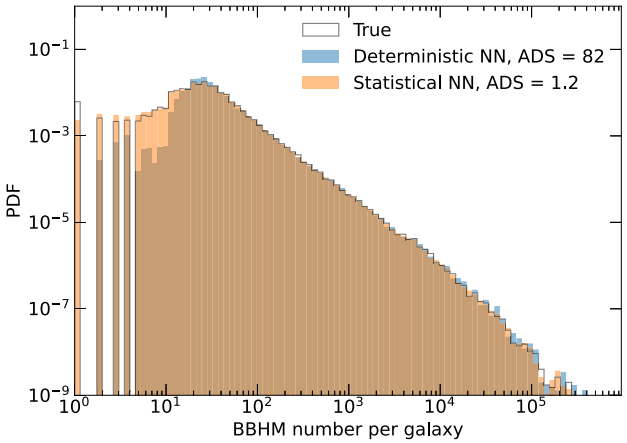

Figure 4. BBHM number PDF per galaxy, marginalized over all the host galaxy and halo properties. The deterministic architecture fails in recovering the 'true' distribution at the low and high number ends of the distribution.

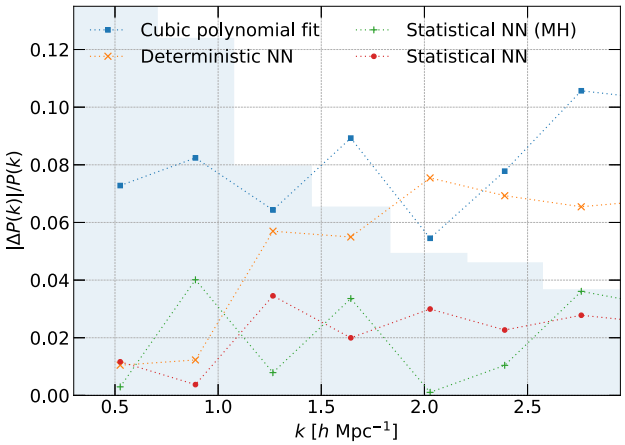

Figure 5. Fractional difference between the N-body power spectrum and different population methods used to paint BBHMs on the simulation. The cubic polynomial fit is already capturing the power spectrum at 10 per cent level. Using the statistical NN approach (red dots) improves the emulation compared to methods not describing the spread of the population (orange x's, namely the deterministic NN). Basing the training on just the host halo mass (green crosses) does not compromise the results.

true sample and the one predicted from the deterministic network are fixed, the samples generated by the statistical network change depending on the random number generator's seed. While the general trends of all the statistical-network-generated samples are consistent between runs, their ADS scores varies. To produce consistent results, we produce 100 realizations, all based on the same galaxy test set, and we concatenate them to calculate the ADS. For the deterministic network, we find ADS = 82, making the significance of the two sets following the same distribution vanishingly small. On the other hand, for the statistical network we calculate ADS = 1.2, thus we cannot discern it from the true distribution at <95 per cent CL.

Secondly, we are interested in analysing the difference in performance at the power spectrum level. In Fig. 5 we show, as a function of wavemodes, the fractional difference between the merger power spectrum extracted from EAGLE25 (our test sample) and the merger power spectrum of the same simulation painted with the different NN architectures. We highlight how (due to the smaller box size) the scales represented here only marginally coincide with those used in our bias analysis, and with the scales relevant for Cosmology. However, in this context, the power spectrum ought to be interpreted as a generic signal that we can use to test our painting scheme, without venturing into any claim about what cosmological information could be extracted from it. While the deterministic NN architecture already achieves a $\lesssim 8$ per cent agreement, which on large scales is negligible with respect to sample variance, the statistical NN architecture leads to a factor $\approx$2 improvement.

Having demonstrated the better performance of the statistical network, we will from now on only use this architecture.

5.3 Feature importance

We test the relative importance of features by re-training the NN with different sets of features, comparing each network performance. Since most of the features are strongly correlated, this is a safer procedure than simply calculating the permutation importance, even if more costly. In the remainder, we use the Spearman's rank correlation coefficient ρ as a measure of correlation (see e.g. Daniel 1990).

We notice that random-forest-based conditional permutation importance (Strobl et al. 2008) would be a valid alternative, that we will consider in future work, to assess every feature separately. Since, on the contrary to the the previous section, the loss function employed by the network is now fixed, we can use as our performance metric not only the ADS, but also the likelihood used as loss function itself. While the ADS is sensitive to the global distribution (regardless of, e.g. host galaxy mass, SFR, metallicity, etc.), the likelihood instead is local in feature space, providing complementary information.

We start by assessing how important the information about the progenitors is. Training the statistical NN on the parameters pertaining to the snapshot where the merger happens, we calculate ADS = 46. This is to be expected (Scelfo et al. 2020; Bellomo et al. 2022) as what is relevant is the physical state of the host at the time of the binary formation. After the formation, the evolution of the binary, until its eventual merging is only slightly influenced by changes in the surrounding environment.

Having assessed that information about the host progenitors is crucial, we turn our attention to five different (groups of) parameters: $M_\star$, $M_{\rm H}$, ψ, Z, and $V_{\rm max}$. In each group, say $M_\star$, we include the value of that parameter for the host, and for the progenitors of the host at $z = 1, 2, 3$, say $[M_\star(z=0), M_\star(z=1), M_\star(z=2), M_\star(z=3)]$. Different determinations of the halo mass are not included, since $M_{\rm H}$ strongly correlates with them at each redshift ($\rho \geq 0.95$), acting as a reliable proxy. The same goes for the parameters related to the kinetic state of the galaxy, namely its internal energy and velocity dispersion, and $V_{\rm max}$ ($|\rho| \geq 0.92$). An exception is made for Z and ψ which are both included despite their correlation ($\rho \geq 0.84$), because of their physical importance and the relevance they usually get in the literature.

We train five networks, each of which only receives as input one single group of variables, and select the best according to ADS and likelihood. In all cases investigated here, the two metrics produced the same rankings. The most relevant feature is fixed and four networks are trained receiving in input that one and another feature, and the process is repeated. The progression is shown in Fig. 6. The host galaxy mass is the most relevant feature to determine the BBHM number distribution (similar to what discussed in A19 and Santoliquido et al. 2021), together with the mass of its first-progenitors at different snapshots. We find that the mass of the host halo and $V_{\rm max}$, which we remind is a proxy of the kinetic state of the galaxy, are also extremely relevant. Deeming this worth of further

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

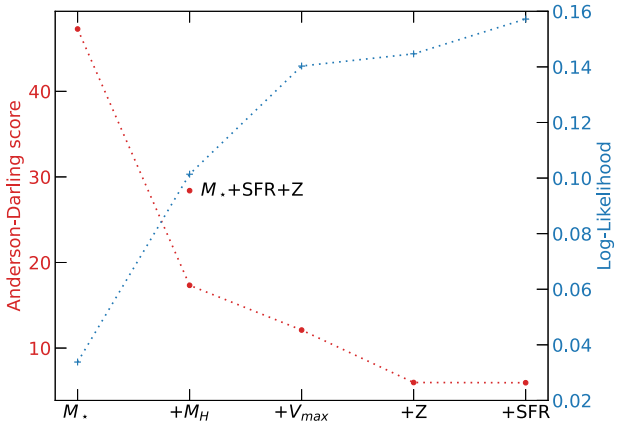

Figure 6. Anderson–Darling score and log-likelihood for different (incremental) sets of features. Each set includes all labels on its left plus the labels indicated on the axis.

investigation we test if this is driven by $M_{\rm H}$, $V_{\rm max}$ and $M_\star$ acting in concert as a proxy of the relevant information contained in ψ and Z. If this were the case, a network trained with ψ, Z, and $M_\star$ would perform better than one trained with only $M_\star$, $M_{\rm H}$, and $V_{\rm max}$. However, as shown in Fig. 6, this does not hold. Therefore, we confirm that the large-scale environment in which the host galaxy forms – meaning its position in the cosmic web, if it is a central or satellite galaxy, etc. – drives the amount of BBHM. This is akin to how the large-scale environment of an halo influences the number and type of galaxies it hosts (Artale et al. 2018).

Z and ψ rank as the least important variables, and the addition of one over the other brings very little additional information. This is not to say that they are physically irrelevant, but rather that since they are extremely correlated to each other and quite correlated to $M_\star$, they turn out to be partially redundant. The link between these galaxy properties is present in the fundamental metallicity relation (see e.g. Mannucci et al. 2010; Curti et al. 2020; Stephenson et al. 2024) which stands for the anticorrelation between Z and ψ, at fixed stellar mass found in observations at local and global scales. This is specifically exploited in, for example, S20 through the use of the AMR.

As for the power spectrum, we find it to be a quite forgiving observable. In fact, even assigning to each galaxy the number of BBHM, using the cubic polynomial fit (i.e. setting all the fractional residual to 1), the error on the power spectrum is $\lesssim 10$ per cent, while training the statistical network just with the host galaxy mass as input, the deviation is $\lesssim 4$ per cent (Fig. 5). Training the network on the full set of parameters does not lead to an improvement. Our analysis highlights a difference in the relative importance of the additional parameters to fully characterize the power spectrum and BBHM distribution. The existence of a relation that adequately links the sole mass of the halo to the BBHM power spectrum, while incredibly useful to populate relatively cheap N-body simulations, is not helpful in understanding the underlying physics. If one's goal is the latter, the use of the BBHM distribution cannot be bypassed. In this sense, our analysis informs how bias models shall be built.

6 CONCLUSIONS

With the fourth LIGO–Virgo–KAGRA observational campaign approaching (Abbott et al. 2020) and third-generation detectors planned for the next decade (Kalogera et al. 2021; Branchesi et al. 2023), it is becoming more and more important to build accurate models to understand the properties of GW sources. These include their redshift distribution and clustering properties, which, in the case of astrophysical binary mergers, ultimately depend on the way their progenitors form according to the properties of their host halo and galaxy.

In this work, we focused on the study of the clustering of astrophysical stellar BBHMs. We relied on cosmological, hydrodynamical simulations populated with BBHMs (McAlpine et al. 2016; A19, A20) to estimate the cosmological linear bias of such events at different redshifts. On one side, we developed a fully numerical estimator (publicly available at: https://github.com/MatPeron/powerbias) which can be applied to mock or observed GW catalogues. Because of the small scales accessible by the size of the simulation adopted, we needed to account for tree-level and one-loop contributions to the power spectrum and to marginalized over nuisance, higher order bias parameters. Our algorithm was tested by comparing the theoretical DM power spectrum and halo bias with their estimates on larger N-body simulations, in which linear scales can be accessed and leading-order truncation can be performed. On the other side, we implemented a semi-analytical HOD-based approach (Libanore et al. 2021, 2022) to estimate the BBHM bias as a function of the properties of their host galaxies. The method was calibrated on the same simulations as the fully numerical estimator and the two approaches show internal consistency. We also compared our results with an independent semi-analytical model (Scelfo et al. 2020) based on merger rate-weighting of the host galaxy bias. Our finding is that the different approaches are consistent, provided that the underlying assumptions on the astrophysical modelling and the properties of the host haloes and galaxies are in agreement. While this seems to be obvious, the important message is that, if we want to use the GW bias as a tool to perform precision cosmological studies, we first of all need to understand how to properly account for uncertainties in the formation channels and host properties. As a first step in this direction, we implemented an NN aimed at studying the relative relevance of the host parameters with respect to the BBHM distribution and bias. While for the former it is important to account for the host galaxy mass, as well as the host halo mass and the galaxy properties at the time of the binary formation, the latter (which in the end depends on how the power spectrum behaves) seems to suppress the relevance of parameters other than the host galaxy or host halo mass.

We note here that our NN can also be applied in principle to populate large sets of N-body, or cosmological hydrodynamic simulations, with BBHM. This could be done very efficiently, since we would no longer have to resort on population synthesis codes after the initial NN training step. While this approach is potentially interesting and worth investigating, it will be crucial to test its robustness in first place. It is for example known (e.g. Villaescusa-Navarro et al. 2021b) that different hydrodynamic models produce qualitatively different simulations; therefore, the network might fail to populate simulations which are not created using the same algorithm that is used to train the model, as seen in, for example, Villaescusa-Navarro et al. (2021a). It must be stressed, however, that this problem is more general and goes beyond the specific choice of relying on a machine-learning trained model to paint merger events on the input galaxy field. This issue would, for example, likely propagate also on catalogues produced using MOBSE, with different hydrodynamic simulations in input. In such scenario, the machine-learning methodology would then still be useful to quickly generate catalogues with self-consistent choices of data and training sets. The problem – in this case – would become that of checking the

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

robustness of summary statistics of GW clustering, extracted from different hydrodynamic models in input – irrespective of how these models are processed (i.e. via machine learning or by repeatedly using population synthesis codes to include the mergers). Another important point is that the NN is trained with fixed cosmological parameters in the input simulation and it would therefore be 'blind' to possible background-cosmology-dependent changes in the merger distribution. We can argue here that, by virtue of a separation of scale argument, the formation of black hole binaries and their subsequent evolution should mostly be influenced by their galactic environment and not by the overall cosmological background. Of course, this hypothesis should also be verified prior to any application. Given their complexity, we postpone a thorough investigation of these aspects to future studies, in which the NN should be tested on simulations accounting for different cosmologies and astrophysical models.

To conclude, with our work we compared the different GW bias prescriptions available on the market and we provided bias estimates based on the state-of-the-art situation in BBHM simulations. Our results in Table 1 and the fit in equation (6) can be used as priors to perform cosmological studies. Improving these findings will be more and more important as the number of GW observations will increase. Once observed GW catalogues will allow to perform statistical studies up to cosmological distances, the clustering will be the perfect tool to perform a wide variety of tests, from the progenitor formation channels [here we only dealt with astrophysical binary black holes, while primordial binary black holes, if they exist, should have a completely different bias (Raccanelli et al. 2016; Scelfo et al. 2020; Libanore et al. 2023)], to modified gravity models (Balaudo et al. 2022; Scelfo et al. 2023) or redshift-space distortion effects (Mukherjee & Wandelt 2018; Mukherjee et al. 2021a). To achieve this goal, the next challenges will be to understand how to access larger scales with the mock catalogues and how to include selection and observational effects, so to realize mock data sets as close as possible to the real ones. Once these will be available, it will be possible to updated the tools, we realized in order to make them applicable to real estimates.

ACKNOWLEDGEMENTS

We acknowledge M. Mapelli's extensive support in realizing the mock data set and interpreting the results. We thank D. Karagiannis, A. Ricciardone, and E. D. Kovetz for useful discussions and the anonymous referee, whose comments helped us improving the readability of the text. We acknowledge support by the project 'Combining Cosmic Microwave Background and Large Scale Structure data: an Integrated Approach for Addressing Fundamental Questions in Cosmology', funded by the MIUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2017 – grant 2017YJYZAH. AR acknowledges support from PRIN-MIUR 2020 METE, under contract no. 2020KB33TP. SL acknowledges support by the Fondazione Ing. Aldo Gini and the Azrieli Foundation. The computational results presented here have been achieved (in part) using the LEO HPC infrastructure of the University of Innsbruck.

DATA AVAILABILITY

The algorithm developed to estimate the (more than tree-level) power spectrum and analyse the clustering properties of EAGLE + MOBSE data set is publicly available at https://github.com/MatPeron/powerbias. The trained NN models are available upon reasonable request. MOBSE is an open source code available at https://gitlab.com/micmap/mobse_open.

REFERENCES

Abbott B. et al., 2016, Phys. Rev. X, 6,041015
Abbott B. P. et al., 2019, Phys. Rev. X, 9, 031040
Abbott B. P. et al., 2020, Living Rev. Relativ., 23, 3
Ade P. A. R. et al., 2014, A&A, 571, A16
Agarwal S., Davé R., Bassett B. A., 2018, MNRAS, 478, 3410
Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, MNRAS, 480, 3978
Artale M. C. et al. 2019, MNRAS, 491, 3419 (A19)
Artale M. C. et al. 2020, MNRAS, 495, 1841 (A20)
Aversa R., Lapi A., de Zotti G., Shankar F., Danese L., 2015, ApJ, 810, 74
Balaudo A., Garoffolo A., Martinelli M., Mukherjee S., Silvestri A., 2022, J. Cosmol. Astropart. Phys., 2023, 42
Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, Phys. Rev. D, 86, 083540
Bellomo N., Bertacca D., Jenkins A. C., Matarrese S., Raccanelli A., Regimbau T., Ricciardone A., Sakellariadou M., 2022, J. Cosmol. Astropart. Phys., 06, 030
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bernardeau F., Colombi S., Gaztanaga E., Scoccimarro R., 2002, Phys. Rept., 367, 1
Boco L., Lapi A., Goswami S., Perrotta F., Baccigalupi C., Danese L., 2019, ApJ, 881, 157 (B19)
Branchesi M. et al., 2023, J. Cosmol. Astropart. Phys., 2023, 191
Calore F., Cuoco A., Regimbau T., Sachdev S., Serpico P. D., 2020, Phys. Rev. Res., 2, 023314
Capurri G., Lapi A., Baccigalupi C., Boco L., Scelfo G., Ronconi T., 2021, J. Cosmol. Astropart. Phys., 11, 032
Chabrier G., 2003, ApJ, 586, L133
Chan K. C., Scoccimarro R., Sheth R. K., 2012, Phys. Rev. D, 85, 083509
Chaniotis A., Poulikakos D., 2004, J. Comput. Phys., 197, 253
Chen S.-F., Vlah Z., White M., 2020, J. Cosmol. Astropart. Phys., 07, 062
Chen S.-F., Vlah Z., Castorina E., White M., 2021, J. Cosmol. Astropart. Phys., 03, 100
Collaboration L. S., Collaboration V., 2021, Phys. Rev. X, 11, 021053
Collaboration T. L. S., the Virgo Collaboration, 2022, preprint (arXiv:2108.01045)
Colombi S., Jaffe A., Novikov D., Pichon C., 2009, MNRAS, 393, 511
Curti M., Mannucci F., Cresci G., Maiolino R., 2020, MNRAS, 491, 944
Daniel W., 1990, Applied Nonparametric Statistics, Duxbury Advanced Series in Statistics and Decision Sciences. PWS-KENT Pub., https://books.google.no/books?id = 0hPvAAAAMAAJ
De Santi N. S. M., Rodrigues N. V. N., Montero-Dorta A. D., Abramo L. R., Tucci B., Artale M. C., 2022, MNRAS, 514, 2463
Dekel A., Lahav O., 1999, ApJ, 520, 24
Del Pozzo W., 2012, Phys. Rev. D, 86, 043011
Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rept., 733, 1
Evans M. et al., 2021, preprint (arXiv:2109.09882)
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Fonseca J., Viljoen J.-A., Maartens R., 2019, J. Cosmol. Astropart. Phys., 12, 028
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publ. Astron. Soc. Pac., 125, 306
Giacobbo N., Mapelli M., Spera M., 2018, MNRAS, 474, 2959
Cañas Herrera G., Contigiani O., Vardanyan V., 2021, ApJ, 918, 20
Howlett C., Lewis A., Hall A., Challinor A., 2012, J. Cosmol. Astropart. Phys., 1204, 027
Iacovelli F., Mancarella M., Foffa S., Maggiore M., 2022, ApJ, 941, 48
Jing Y. P., 2005, ApJ, 620, 559
Jung G. et al. 2022, ApJ, 940, 17
Kalogera V. et al., 2021, preprint (arXiv:2111.06990)
Kamdar H. M., Turk M. J., Brunner R. J., 2016, MNRAS, 457, 1162
Karagiannis D., Lazanu A., Liguori M., Raccanelli A., Bartolo N., Verde L., 2018, MNRAS, 478, 1341

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

Kingma D. P., Ba J., 2017, preprint (arXiv:1412.6980)
Klambauer G., Unterthiner T., Mayr A., Hochreiter S., 2017, preprint (arXiv:1706.02515)
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Li Y., Hu W., Takada M., 2016, Phys. Rev. D, 93, 063507
Libanore S. et al., 2021, JCAP, 02, 035
Libanore S., Artale M. C., Karagiannis D., Liguori M., Bartolo N., Bouffanais Y., Mapelli M., Matarrese S., 2022, JCAP, 02, 003
Libanore S., Liguori M., Raccanelli A., 2023, JCAP, 08, 055
McDonald P., Roy A., 2009, J. Cosmol. Astropart. Phys., 2009, 020
Maggiore M. et al., 2020, J. Cosmol. Astropart. Phys., 2020, 050
Mancuso C., Lapi A., Shi J., Gonzalez-Nuevo J., Aversa R., Danese L., 2016, ApJ, 823, 128
Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115
Mapelli M., Giacobbo N., 2018, MNRAS, 479, 4391
Mapelli M., Giacobbo N., Ripamonti E., Spera M., 2017, MNRAS, 472, 2422
Mapelli M., Giacobbo N., Toffano M., Ripamonti E., Bressan A., Spera M., Branchesi M., 2018, MNRAS, 481, 5324
Marsaglia G., 2006, J. Stat. Softw., 16, 1
Matarrese S., Verde L., Heavens A. F., 1997, MNRAS, 290, 651
McAlpine S. et al., 2016, Astron. Comput., 15, 72
Mead A., Peacock J., 2014, MNRAS, 440, 1233
Mead A. J., Brieden S., Tröster T., Heymans C., 2021, MNRAS, 502, 1401
Mohammed I., Seljak U., Vlah Z., 2017, MNRAS, 466, 780
Mukherjee S., Silk J., 2020, MNRAS, 491, 4690
Mukherjee S., Wandelt B. D., 2018, preprint (arXiv:1808.06615)
Mukherjee S., Wandelt B. D., Nissanke S. M., Silvestri A., 2021a, Phys. Rev. D, 103, 043520
Mukherjee S., Lavaux G., Bouchet F. R., Jasche J., Wandelt B. D., Nissanke S. M., Leclercq F., Hotokezaka K., 2021b, A&A, 646, A65
Mukherjee S., Krolewski A., Wandelt B. D., Silk J., 2022, preprint (arXiv:2203.03643)
Namikawa T., 2021, J. Cosmol. Astropart. Phys., 2021, 036
Namikawa T., Nishizawa A., Taruya A., 2016, Phys. Rev. Lett., 116, 121302
Nix D., Weigend A., 1994, in Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), Vol. 1. IEEE, Orlando, FL, USA, p. 55
Oguri M., 2016, Phys. Rev. D, 93, 083511
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Peebles P. J., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton
Punturo M. et al. 2010, Class. Quantum Gravity, 27, 194002
Raccanelli A., Kovetz E. D., Bird S., Cholis I., Muñoz J. B., 2016, Phys. Rev. D, 94,023516
Santoliquido F., Mapelli M., Giacobbo N., Bouffanais Y., Artale M. C., 2021, MNRAS, 502, 4877
Santoliquido F., Mapelli M., Artale M. C., Boco L., 2022, MNRAS, 516, 3297
Sathyaprakash B. et al., 2012, Class. Quantum Gravity, 29, 124013
Scelfo G., Bellomo N., Raccanelli A., Matarrese S., Verde L., 2018, J. Cosmol. Astropart. Phys., 2018, 039
Scelfo G., Boco L., Lapi A., Viel M., 2020, J. Cosmol. Astropart. Phys., 10, 045 (S20)
Scelfo G., Spinelli M., Raccanelli A., Boco L., Lapi A., Viel M., 2022, J. Cosmol. Astropart. Phys., 2022, 004
Scelfo G., Berti M., Silvestri A., Viel M., 2023, J. Cosmol. Astropart. Phys., 02, 010
Scholz F. W., Stephens M. A., 1987, J. Am. Stat. Assoc., 82, 918
Scoccimarro R., Zaldarriaga M., Hui L., 1999, ApJ, 527, 1
Schaye J. et al. 2015, MNRAS, 446, 521
Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, MNRAS, 460, 3624
Smith L. N., 2017, preprint (arXiv:1506.01186)
Smith R. E. et al., 2003, MNRAS, 341, 1311
Spera M., Mapelli M., 2017, MNRAS, 470, 4739
Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, J. Mach. Learn. Res., 15, 1929
Stephenson H. M. O. et al., 2024, MNRAS, 527, 7891
Stiskalek R., Bartlett D. J., Desmond H., Anbajagane D., 2022, MNRAS, 514, 4026
Strobl C., Boulesteix A.-L., Kneib T., Augustin T., Zeileis A., 2008, BMC Bioinform., 9, 307
Tegmark M., Hamilton A. J. S., Strauss M. A., Vogeley M. S., Szalay A. S., 1998, ApJ, 499, 555
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlober S., 2010, ApJ, 724, 878
Vijaykumar A., Saketh M. V. S., Kumar S., Ajith P., Choudhury T. R., 2023, Phys. Rev. D, 108, 103017
Villaescusa-Navarro F. et al., 2020, ApJS, 250, 2
Villaescusa-Navarro F. et al., 2021a, preprint (arXiv:2109.09747)
Villaescusa-Navarro F. et al., 2021b, ApJ, 915, 71
Wagner C., Schmidt F., Chiang C.-T., Komatsu E., 2015, MNRAS, 448, L11
Yang Q., Hu B., 2022, Phys. Dark Universe, 40, 101206
Zhu L.-G., Hu Y.-M., Wang H.-T., Zhang J.-d., Li X.-D., Hendry M., Mei J., 2022, Phys. Rev. Res., 4, 013247

APPENDIX A: ONE-LOOP BIAS OPERATORS

In equations (2) and (3), we define the autopower and cross-power spectrum expansions adopted in the EFTofLSS estimator. The tree-level and one-loop contributions that enter such equations are defined as in Desjacques et al. (2018) and presented here for ease of reference:

$$
\begin{aligned}
F_{\delta^2}(k) &\equiv 0\,,\\
I_{\delta^2}(k) &\equiv 2\int_{\boldsymbol{q}} F_2(\boldsymbol{k}-\boldsymbol{q},\,\boldsymbol{k})P_{\rm L}(k)P_{\rm L}(|\boldsymbol{k}-\boldsymbol{q}|)\,,\\
F_{K^2}(k) &\equiv 4P_L(k)\int_{\boldsymbol{q}} \sigma^2_{q,\,k-q}F_2(\boldsymbol{k},\,-\boldsymbol{q})P_{\rm L}(q)\,,\\
I_{K^2}(k) &\equiv 2\int_{\boldsymbol{q}} \left(\sigma^2_{q,\,k-q}+\frac{2}{3}\right)F_2(\boldsymbol{k}-\boldsymbol{q},\,\boldsymbol{k})P_{\rm L}(k)P_{\rm L}(|\boldsymbol{k}-\boldsymbol{q}|)\,,\\
F_{\rm td}(k) &\equiv \frac{8}{5}P_L(k)\int_{\boldsymbol{q}} \sigma^2_{q,\,k-q}F_2(\boldsymbol{k},\,-\boldsymbol{q})P_{\rm L}(q)\,,\\
I_{\rm td}(k) &\equiv 0\,,\\
I_{\delta^2\delta^2}(k) &\equiv 2\int_{\boldsymbol{q}} P_L(q)P_{\rm L}(|\boldsymbol{k}-\boldsymbol{q}|)\,,\\
I_{\delta^2K^2}(k) &\equiv 2\int_{\boldsymbol{q}} \left(\sigma^2_{q,\,k-q}+\frac{2}{3}\right)P_{\rm L}(k)P_{\rm L}(|\boldsymbol{k}-\boldsymbol{q}|)\,,\\
I_{K^2K^2}(k) &\equiv 2\int_{\boldsymbol{q}} \left(\sigma^2_{q,\,k-q}+\frac{2}{3}\right)^2P_L(k)P_{\rm L}(|\boldsymbol{k}-\boldsymbol{q}|)\,,
\end{aligned}
$$

where $b_{\delta^2}\equiv b_{2,t}/2$, $\sigma^2_{q,\,k}=(\boldsymbol{q}\cdot\boldsymbol{k}/qk)^2-1$, $P_{\rm L}$ is the linear matter power spectrum, and F_2 is the standard second-order kernel for Einstein–deSitter cosmology.

APPENDIX B: POWER SPECTRUM ESTIMATOR

The snapshots and catalogues we used in our analysis contain $N_{\rm p}$ point-like particles (DM or tracer) distributed inside a cubic box of size L. Each particle is localized within the box by $\boldsymbol{r}=(x,\,y,\,z)$, where the origin $(0,\,0,\,0)$ is at one of the box vertices. We define the density as $\rho(\boldsymbol{r})\equiv m\sum_{i=1}^{N_{\rm p}}\delta_D(\boldsymbol{r}-\boldsymbol{r}_i)$ and the density contrast as

$$
\delta(\boldsymbol{r})\equiv\frac{\rho(\boldsymbol{r})}{\bar{\rho}}-1=\frac{\sum_{i=1}^{N_{\rm p}}\delta_D(\boldsymbol{r}-\boldsymbol{r}_i)}{\bar{n}}-1, \tag{B1}
$$

where $\bar{\rho}=m\bar{n}=mN_{\rm p}/L^3$ is the spatially averaged density within the box. On a theoretical level, the power spectrum is then defined by taking the Fourier transform $\delta(\boldsymbol{k})$, computing its square modulus

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

$|\delta(\boldsymbol{k})|^2 = \delta(\boldsymbol{k})\delta^*(\boldsymbol{k})$, and averaging the resulting modes inside k bins. This procedure assumes that the fair sample hypothesis holds.

Since we are dealing with discrete data sets, the density field is unevenly sampled in space. Therefore, before applying the FFT we need to smear $\delta(\boldsymbol{r})$ on to a regular grid within the box by convolving it with a proper filter (Tegmark et al. 1998). In this way, we interpolate $\tilde{\delta}(\boldsymbol{r})$; to recover $\delta(\boldsymbol{k})$, one then has to deconvolve the filter, which in Fourier space is simply done by dividing the Fourier transform of the filter to $\tilde{\delta}(\boldsymbol{k})$.

With this basic recipe in mind, we can discuss in more detail the steps we take in order to arrive to an estimator $\tilde{P}(k)$.

(i) The filter function $W(\boldsymbol{r})$ we adopt is the standard CIC weighting scheme (see equation A.2 of Chaniotis & Poulikakos 2004), which smears a single particle across a cube of eight neighbouring cells. Chosen a number of grid cells per side $N_{\rm g}$, we use it to compute the weights $W(\boldsymbol{r}_{\rm g})$ for each particle at each grid's node position $\boldsymbol{r}_{\rm g}$. The weights are then summed across the entire catalogue to estimate the number density field $\tilde{n}(\boldsymbol{r}_{\rm g}) = \tilde{\rho}(\boldsymbol{r}_{\rm g})/m$.

(ii) $\tilde{\delta}(\boldsymbol{r}_{\rm g})$ is found according to equation (B1).

(iii) $\tilde{\delta}(\boldsymbol{k})$ is given by performing the FFT of $\tilde{\delta}(\boldsymbol{r}_{\rm g})$.

(iv) The Fourier transform of the CIC filter function $W(\boldsymbol{r})$ is computed analytically and divided to $\tilde{\delta}(\boldsymbol{k})$ in order to recover the actual Fourier transform of the field, $\delta(\boldsymbol{k})$.

(v) The proper normalizations given by the FFT and the discretization of the Dirac deltas are applied to $\delta(\boldsymbol{k})$.

(vi) We compute $\tilde{P}(k)$ by averaging $|\delta(\boldsymbol{k})|^2$ inside bins in Fourier space. Bins are defined inside the interval $[k_{\rm f},\ k_{\rm Nyq})$ and are $k_{\rm f}$-wide, where $k_{\rm f} = 2\pi/L$ is the fundamental mode, while $k_{\rm Nyq} = \pi N_{\rm g}/L$ is the Nyquist mode, i.e. the maximum mode supported by the grid.

(vii) We compute error bars for the estimator by taking the square root of the Gaussian variance, $\sqrt{\mathrm{Var}(\tilde{P}(k))|_{\rm g}} = \sqrt{2/N_k}\,\tilde{P}(k)$, where N_k is the number of independent Fourier modes within the bin (Scoccimarro et al. 1999). We neglect non-Gaussian contributions related with the four-point function of $\delta(\boldsymbol{k})$ (namely, the trispectrum) since they are only significant at even smaller scales than the ones we consider (Mohammed, Seljak & Vlah 2017).

Since we are computing $\tilde{P}(k)$ from a discrete collection of particles, shot-noise arises from their autocorrelations. We can assume that the particles are Poisson sampled from the underlying density field; under this assumption, the shot noise contribution is equal to $\bar{n}^{-1}$ (see appendix A of Feldman et al. 1994). In our analysis, we subtract it before making any measurement of the bias parameters.

Another consideration to make is that by choosing to smear the particles on to a regular grid, we are effectively shifting all the power associated to the modes $k > k_{\rm Nyq}$ inside the range of modes supported by the grid. This effect is known as aliasing and it causes a spurious increase in power at modes $k \gtrsim k_{\rm Nyq}/2$ (Colombi et al. 2009). Aliasing can be alleviated (Sefusatti et al. 2016) or outright removed under certain assumptions on the shape of $P(k)$ (Jing 2005). We implemented the interlacing approach from Sefusatti et al. (2016) to deal with this effect (see Fig. C1). It is important to note that interlacing, while useful if seeking accurate estimation from large-scale simulations, is not necessary for small-side simulations such as EAGLE. In fact, the range of modes where the perturbative prescription of equation (1) is valid is very close to the fundamental mode, where aliasing has no effect. For this reason, we only use interlacing to analyse QUIJOTE data in the validation test performed in Appendix C.

The procedure we depicted up to now is optimal to estimate the power spectrum on ideal, mock data sets; with the prospect of applying the same methodology to real data, however, selection effects should be taken into account. For this reason, in our algorithm, we also included the FKP estimator (Feldman et al. 1994), in which the data set is compared with a synthetic random catalogue, in order to optimize the variance. In our case, this catalogue is generated by uniform sampling a set of particles inside the same box in which the simulation has been run. The implementation follows the same general scheme presented above, with some extra-steps related with the synthetic catalogue and the optimal weighting that characterizes the FKP estimator. In particular, step (i) in our algorithm is applied twice, once on the mock catalogue and another on the synthetic, random catalogue. Then, before step (ii), we define the optimal weights $w(\boldsymbol{r}) = [\bar{n}(\boldsymbol{r})P_0(k) + 1]^{-1}$ from a initial guess of the power spectrum $P_0(k) = \mathrm{const}$. These are used to compute the weighted number density field $F(\boldsymbol{r})$:

$$F(\boldsymbol{r}) = \frac{w(\boldsymbol{r})\,(n(\boldsymbol{r}) - \alpha n_s(\boldsymbol{r}))}{\sqrt{\int \mathrm{d}^3 r\; w(\boldsymbol{r})^2 \bar{n}(\boldsymbol{r})^2}}, \tag{B2}$$

where the s subscript refers to the synthetic random catalogue, while $\alpha \lesssim 0.1$ is the ratio between the density of the data catalogue and the synthetic, random one. The algorithm then proceeds using $F(\boldsymbol{r})$ in place of $\delta(\boldsymbol{r})$. The presence of $P_0(\boldsymbol{k})$ in $w(\boldsymbol{r})$ implies that the weights can be optimized iteratively by using the resulting $\tilde{P}_{\rm FKP}(k)$. Therefore, by starting from $P_0(k) = \mathrm{const}$, we get a first estimate $\tilde{P}_{\rm FKP}(k)$ which is used to improve the estimate of the optimal weights bin-by-bin; for each k, we then compute a new set of weights $w(\boldsymbol{r})$ using $P_0(k) = \tilde{P}_{\rm FKP}(k)$, and obtain a new estimate $\tilde{P}_{\rm FKP}(k)$. This is iterated up to some convergence criterion.

APPENDIX C: VALIDATION TESTS

As we anticipated in Section 2, the small volume of the EAGLE simulation (Schaye et al. 2015) forces us to estimate the power spectrum and bias on small, non-linear scales. Theoretical models in this regime are complicated and depend on large sets of parameters. Before using our algorithm on these scales, therefore, we tested it in a 'simpler' configuration, namely on the large, linear scales provided by the QUIJOTE simulation (Villaescusa-Navarro et al. 2020).

We now describe the validation tests we performed both on the power spectrum and bias estimators.

C1 Power spectrum estimator

We validate our power spectrum estimator on the DM field from the QUIJOTE simulations (Villaescusa-Navarro et al. 2020). We consider five different realizations with the same cosmological parameters (the fiducial ones). For each, we compute the DM power spectrum for every available redshift snapshot, namely $z \in \{0,\ 0.5,\ 1,\ 2,\ 3\}$, interpolating its density contrast field on to a grid of $N_{\rm g} = 512$ cells per side. The resulting 25 power spectra are then averaged across the five realization, obtaining five estimated power spectra, one for each redshift. We then use CAMB[6] (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012) to compute the theoretical non-linear matter power spectrum given by the halofit model (Mead et al. 2021) for the given cosmology at each redshift.

Fig. C1 compares the estimated and theoretical power spectra at $z = 0$. Up to $k \sim 0.2\ h\,\mathrm{Mpc}^{-1}$, the power spectrum measured from the simulations agrees within 5 per cent with the theoretical one. At smaller scales a systematic shift in power is observed, which can be

[6] https://github.com/cmbant/CAMB.

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

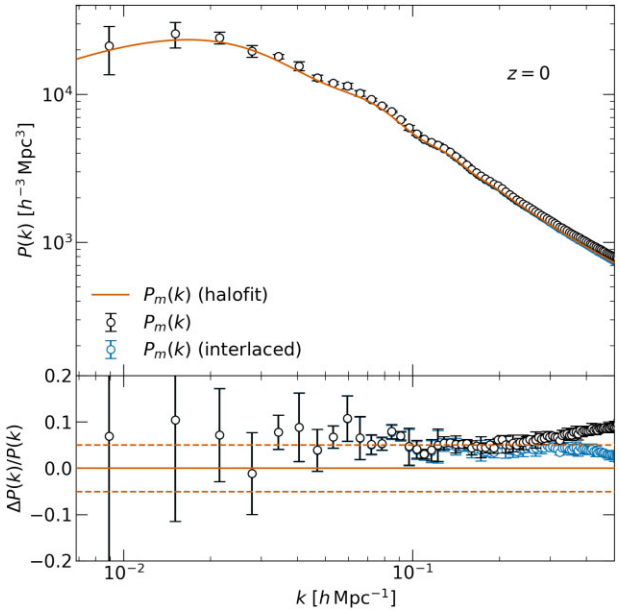

Figure C1. QUIJOTE DM power spectrum estimated at $z = 0$ with (blue) and without (black) interlacing, compared with $P(k)$ from CAMB (orange line).

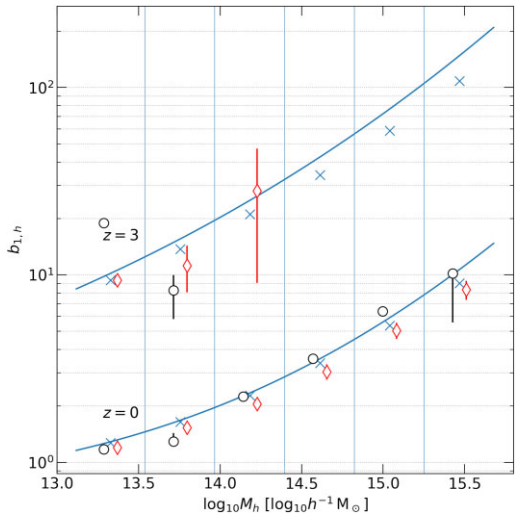

Figure C2. Halo bias at $z = 0$ and 3 estimated with MCMC based on EFTOFLSS (black circles), compared with the PBS expectation value weighted by the halo mass function (blue crosses) and the empirical linear bias estimator (red diamonds).

explained by aliasing (Jing 2005; Sefusatti et al. 2016). This behavior is observed at all redshifts. Using the interlacing algorithm introduced in Sefusatti et al. (2016) improves the estimate considerably, pushing the agreement up to $k \sim 0.4{-}0.5\ h\,\mathrm{Mpc}^{-1}$.

C2 Bias estimator

We test the MCMC estimator described in Section 4.1 on DM haloes from QUIJOTE simulations (Villaescusa-Navarro et al. 2020).

We consider the same five realizations used in the previous section, sorting the haloes into six mass bins uniform in log-space. For each mass bin and redshift, we compute the interlaced power spectrum using a grid of $N_g = 512$ cells per side to interpolate the density field. We obtain a total of 118 power spectra (we are unable to

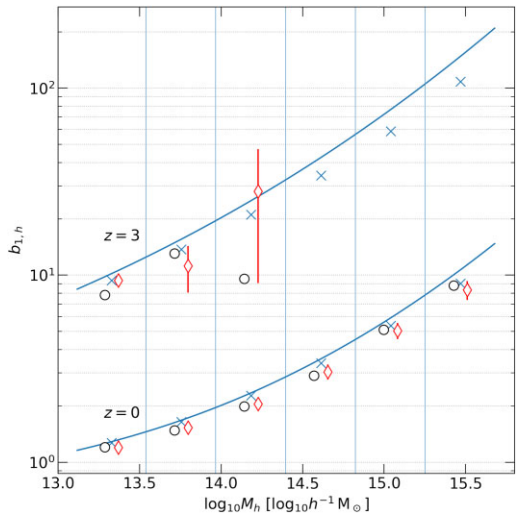

Figure C3. Halo bias at $z = 0$ and 3 estimated with MCMC based on LIMD (black circles), compared with the PBS expectation value weighted by the halo mass function (blue crosses) and the empirical linear bias estimator (red diamonds).

compute 32 of them due to the absence of haloes at high redshifts and masses). Again, we average the different realizations, obtaining 24 estimated power spectra, one for each redshift and mass bin.

Finally, we obtain estimates for the linear bias parameter $b_{1,\mathrm{h}}$ by fitting the models of equations (2)–(5) to the estimated power spectra. As in Section 4.1, we use a Gaussian likelihood and uniform priors, taking care in setting $b_{1,\mathrm{h}} \in [0,\ 50]$ to provide enough parameter space for the chains to explore the posterior. We combine auto and cross-power spectra corrected for the shot noise and truncated at a scale k_{NL}, which we set at 90 per cent the non-linear threshold scale for matter described in Smith et al. (2003); Fonseca, Viljoen & Maartens (2019), that is $k_{\mathrm{NL}} = 0.9 \cdot 0.2(z+1)^{2/(2+n_s)}\ h\,\mathrm{Mpc}^{-1}$ with $n_s = 0.9624$.

We compare these estimates with the theoretical expectations of $b_{1,\mathrm{h}}$ from the PBS method, assuming the theoretical halo mass function distribution described by Tinker et al. (2010) and computed through the `hmf` python package. The PBS gives us $b_{1,\mathrm{h}}(M_{\mathrm{h}})$ as a function of the halo mass M_{h}; we use the halo mass function to weigh it inside each mass bin, so to find a single value per bin. Figs C2 and C3 compare the EFTOFLSS and LIMD halo bias estimates with the PBS expectation values weighted in each mass bin. Using the Tinker halo mass function to weigh, the PBS bias gives results that are in overall agreement with the empirical estimations at linear scales, with some small deviations due to differences between the actual halo mass function of the simulation compared to the Tinker prescription.

As a further test, we estimate the linear bias truncating the power spectra at $k \sim 0.08\ h\,\mathrm{Mpc}^{-1}$. This leaves us only with the linear scales, where the bias expansions can be truncated at leading order:

$$P_{\mathrm{tt}}(k) = b_{1,\mathrm{t}}^2 P_{\mathrm{L}}(k) + P_\epsilon, \tag{C1}$$

$$P_{\mathrm{td}M}(k) = b_{1,\mathrm{t}} P_{\mathrm{L}}(k). \tag{C2}$$

$b_{1,\mathrm{h}}$ can easily be obtained by reversing these equations. These estimates are then averaged together across the various k-modes to obtain a final estimate of $b_{1,\mathrm{h}}$, with 1σ error bars given by their standard deviation. This empirical bias estimate is compared with the MCMC EFTOFLSS and LIMD results in Figs C2 and C3.

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025

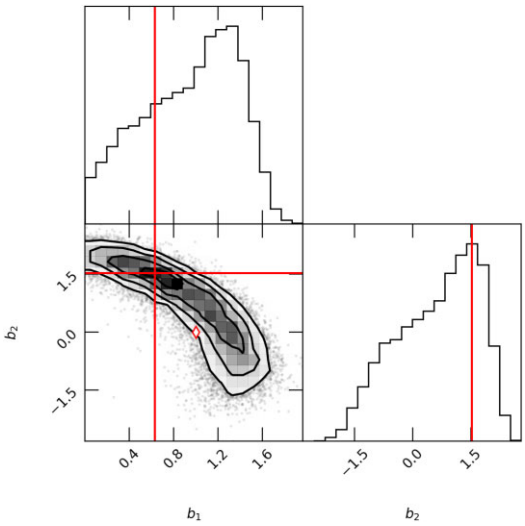

Figure C4. Posterior and marginalized distributions of b_1 and b_2 obtained through a joint MCMC sampling of the autopower spectrum of BBHMs, and cross-power spectrum between mergers and matter computed at redshift $z = 0$. The red lines indicate the maximum a posteriori estimate of the parameters, while the red diamond indicates the initial guess.

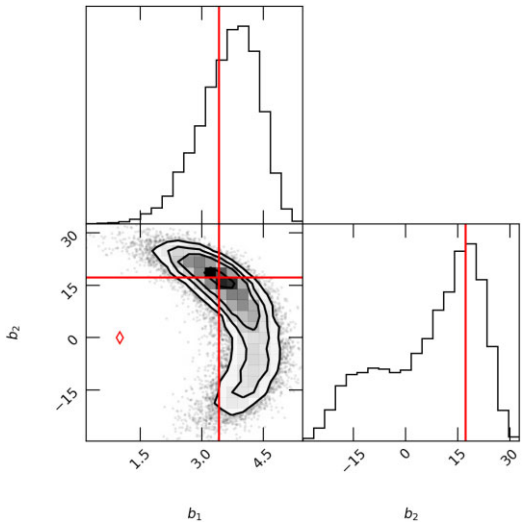

Figure C5. Same as C4 at $z = 6$.

MCMC estimates obtained with the LIMD model agree really well with the empirical ones, with some noise at high redshifts explainable by the small amount of haloes in this range. On the other hand, EFTofLSS results while following the same trend as the other estimator, are noisier, which may be explained by the higher dimensionality of the model itself. It is important to note that, while these results favour the LIMD model over the EFTofLSS one, the BBHM catalogues on which we recover the final result of this paper contain far more objects than the DM haloes used here, meaning that the measured power spectra will be less noisy. Indeed, Section 4.1 shows that LIMD and EFTofLSS are in perfect agreement when it comes to BBHMs.

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

© 2024 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mnras/article/530/1/1129/7638208 by Universita degli studi di Padova - Dipartimento di diritto pubblico internazionale e comunitario user on 10 February 2025