Extracting cosmological information from the abundance of galaxy clusters with simulation-based inference

Íñigo Zubeldia,[1,2,*] Boris Bolliet,[3,2] Anthony Challinor,[1,2,4] and William Handley[1,2]

[1]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*

[2]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*

[3]*Cavendish Astrophysics, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*

[4]*DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom*

(Received 14 April 2025; accepted 18 September 2025; published 17 October 2025)

The abundance of galaxy clusters as a function of mass and redshift is a well-established and powerful cosmological probe. Cosmological analyses based on galaxy cluster number counts have traditionally relied on explicitly computed likelihoods, which are often challenging to develop with the required accuracy and expensive to evaluate. In this work, we implement an alternative approach based on simulation-based inference (SBI) methods that relies solely on synthetic galaxy cluster catalogs generated under a given model. These catalogs are much easier to produce than it is to develop and validate a likelihood. We validate this approach in the context of the galaxy cluster survey of the upcoming Simons Observatory for a setup in which we can also evaluate an exact explicit likelihood. We find that our SBI-based approach yields cosmological parameter posterior means that are within 0.2σ of those obtained with the explicit likelihood and with biases smaller than 0.1σ. We also introduce and validate a procedure to assess the goodness of fit using only synthetic catalogs similar to those used for training. This demonstrates, for the first time, that a galaxy cluster number count cosmological analysis can be performed fully without resorting to a likelihood at any stage. Finally, we apply our SBI-based approach to the real *Planck* MMF3 cosmology sample, obtaining cosmological parameter constraints that are within 0.1σ of their likelihood-based counterparts. This constitutes the first SBI-based number count cosmological analysis of a real galaxy cluster catalog.

DOI: 10.1103/drn7-ggqk

I. INTRODUCTION

The abundance of galaxy clusters as a function of mass and redshift is a powerful cosmological probe, sensitive to cosmological parameters such as the matter density parameter $\Omega_{\rm m}$, the amplitude of matter clustering σ_8, the equation of state of dark energy, and the sum of the neutrino masses ([1–4]). This has been demonstrated over the past three decades in a number of cosmological "cluster number count" analyses using clusters detected in X-ray, optical and millimeter observations (e.g., [5–31]). With 10^4–10^5 objects, cluster catalogs from current and upcoming observatories such as eROSITA [32], *Euclid* [33], the Vera C. Rubin Observatory (Rubin/LSST; [34]), SPT-3G [35], the Simons Observatory (SO; [36]), and CMB-S4 [37] have the potential to improve the constraints derived from their predecessors significantly, taking advantage of their sheer statistical power and of the synergies between different observations (e.g., using galaxy weak-lensing observations to calibrate X-ray and thermal Sunyaev–Zeldovich (tSZ) mass-observable scaling relations empirically, as in, e.g., [14,27,28], or combining cluster counts with cluster clustering, as in, e.g., [20]).

Thus far, all published cosmological cluster number count analyses have relied on explicitly computing the likelihood of the data. Cluster number count likelihoods are complex objects in which the population of clusters across redshift and one or more mass observables is typically modeled with a Bayesian hierarchical population model. Evaluating these likelihoods often requires the efficient computation of numerous multidimensional integrals (see, e.g., [12,14,15,28,38,39]). Developing a likelihood for a given cluster catalog, and ensuring its required accuracy

*Contact author: inigo.zubeldia@ast.cam.ac.uk

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

across the relevant region of parameter space with the constraint of an acceptable evaluation time, have traditionally demanded a very significant investment of resources.

In this work we implement an alternative approach for obtaining cosmological constraints from galaxy cluster catalogs that does not require the explicit use of a likelihood. This approach makes use of recent developments in simulation-based inference (SBI) methods (see [40] for a review), in particular in neural posterior estimation, and relies solely on synthetic cluster catalogs generated under a given model. These catalogs are significantly easier to generate than it is to develop and validate an explicit likelihood, making this SBI-based approach particularly attractive. We validate our approach in the context of an SO-like survey, but expect it to be broadly applicable. We then apply it to a real dataset, the *Planck* MMF3 cosmology galaxy cluster sample [13,41].

This paper is organized as follows. First, in Sec. II we describe our SBI-based approach and validate it for an SO-like survey, also discussing the goodness of fit within the SBI framework. Next, in Sec. III we apply it to the *Planck* MMF3 cosmology sample and in Sec. IV we compare our results to previous work. We then discuss the advantages of our SBI-based approach in Sec. V before concluding in Sec. VI. Appendix A lists the parameter priors and the constraints that we obtain in our analyses, and Appendix B discusses the run-time performance of our SBI-based approach relative to the traditional likelihood-based one.

II. SIMULATION-BASED INFERENCE FOR GALAXY CLUSTER COSMOLOGY

A. Demonstration scenario: The Simons Observatory cluster survey

In a galaxy cluster number count cosmological analysis, the input dataset consists of a catalog of galaxy clusters. Each cluster must have at least one of the following: (i) a redshift measurement; and (ii) measurements for one or several mass observables, i.e., cluster observables that scale with cluster mass.

In this work, we test our proposed cosmological inference method in the context of the galaxy cluster catalog due to be delivered by the Simons Observatory, an upcoming state-of-the-art millimeter observatory. One of the major goals of the SO Collaboration's cosmological program is to produce a catalog of galaxy clusters detected via the tSZ effect and to derive number count cosmological constraints from it. For SO's baseline noise levels and assuming a cosmology consistent with *Planck* CMB constraints, about 16 000 galaxy clusters are forecast to be detected [36,39], with this number expected to double for the full Advanced SO experiment [42].

Here, we consider an SO-like cluster survey covering 40% of the sky with a catalog in which every cluster has a measurement for two mass observables: the tSZ signal-to-noise, $q_{\rm obs}$, and the CMB lensing signal-to-noise, $p_{\rm obs}$. Every cluster also has a redshift measurement, z, yielding a total of three data points per cluster. The catalog is assumed to be constructed by imposing a selection threshold of $q_{\rm obs} = 5$. The cosmological information is extracted from the cluster abundance in the $q_{\rm obs}$–z plane, with the $p_{\rm obs}$ measurements serving to calibrate the scaling relation between cluster mass and the tSZ signal-to-noise, which cannot be accurately predicted from first principles.

The two mass observables are linked to cluster mass and redshift with a two-layer hierarchical model. In the first layer, the *mean* tSZ signal-to-noise $\bar{q}(M_{500}, z)$ is given by

$$\bar{q}(M_{500}, z) = \frac{y_0(\beta_{\rm SZ} M_{500}, z)}{\sigma_{y_0}(\theta_{500}(\beta_{\rm SZ} M_{500}, z))}, \tag{1}$$

where y_0 is the cluster's central Compton-y value, σ_{y_0} is the cluster detection noise evaluated at the cluster's angular scale θ_{500}, M_{500} is the cluster mass, and the tSZ mass bias is $\beta_{\rm SZ} = 0.8$. We relate y_0 to mass by

$$y_0 = 10^{A_{\rm SZ}} \left(\frac{M_{500}}{3 \times 10^{14} h_{70}^{-1} M_\odot} \right)^{\alpha_{\rm SZ}} E^2(z) h_{70}^{-1/2}, \tag{2}$$

where $A_{\rm SZ}$ parametrizes the tSZ signal amplitude and $\alpha_{\rm SZ}$ its mass dependence. Furthermore, $E(z)$ is the Hubble parameter scaled by the current value, $H_0 = 100\, h\, {\rm km\, s^{-1}\, Mpc^{-1}}$, and $h_{70} = h/0.7$. This scaling relation is consistent with the universal pressure profile of [43] and its form follows the tSZ scaling relation of [44]. The detection noise σ_{y_0} is computed by applying the tSZ cluster finder SZIFI[1] [45,46] to SO-like maps from the Websky simulation [47,48]; see [39], where similar catalogs are used, for further details.

We note that fixing the value of $\beta_{\rm SZ}$ does not reduce the generality of our analysis, as $A_{\rm SZ}$ and $\alpha_{\rm SZ}$ are empirically calibrated with the CMB lensing measurements, with wide uniform priors imposed on them (see below); fixing $\beta_{\rm SZ}$ can simply be interpreted as modifying the pivot mass of $3 \times 10^{14} M_\odot$. Since we assume true values for $A_{\rm SZ}$ and $\alpha_{\rm SZ}$ that are derived from X-ray observations (specifically, from [43]; see Table I for their values), we have introduced $\beta_{\rm SZ}$ and set its value to $\beta_{\rm SZ} = 0.8$. This value is consistent with the relatively broad range of constraints derived from data and hydrodynamical simulations over the past decade (e.g., [15,49–51]).

Still in the first layer of the hierarchical model, the *mean* CMB lensing signal to noise, $\bar{p}(M_{500}, z)$, is given by

$$\bar{p}(M_{500}, z) = \frac{\kappa_0(\beta_{\rm CMBlens} M_{500}, z)}{\sigma_{\kappa_0}(\theta_{500}(\beta_{\rm CMBlens} M_{500}, z))}, \tag{3}$$

[1] github.com/inigozubeldia/szifi

where $\kappa_0(M_{500}, z)$ is the central value of the cluster's CMB lensing convergence, σ_{κ_0} is the CMB lensing matched-filter noise, and the CMB lensing mass bias is $\beta_{\rm CMBlens} = 0.92$ (a value chosen to be consistent with the simulation-based calibration of [52]). This expression assumes that the cluster CMB lensing signal has been extracted with a matched-filter approach, as first proposed in [53] and applied, e.g., in [13,15,52,54]. Following [15,52], in order to compute both κ_0 and the matched-filter noise, we assume that the cluster convergence profile is that of a truncated Navarro-Frenk-White profile ([55]) with a concentration $c_{500} = 3$ and a truncation radius of $5R_{500}$, with $\beta_{\rm CMBlens}$ accounting for the bias in the inferred masses due to the mismatch between the assumed cluster profile and the true mean one (see [52]). We compute the matched-filter noise using the publicly available SO minimum-variance (temperature + polarization) quadratic estimator reconstruction noise curve forecast,[2] taking it to be the same for all the clusters in the sample.

The logarithms of the mean tSZ and CMB lensing signals to noise are then linked to the logarithms of the *true* tSZ and CMB lensing signals to noise through Gaussian intrinsic scatter, with a covariance matrix given by $\sigma_{\rm SZ}$, $\sigma_{\rm CMBlens}$, and a correlation coefficient $r = 0$ (i.e., no intrinsic correlation). This log-normal intrinsic scatter accounts for true scatter in the cluster observables (e.g., due to cluster profile variations, triaxiality, and large-scale structure along the line of sight). In the second layer of the model, the set of scaling relations simply exponentiates $\ln q$ and $\ln p$, which are then linked to the observed values ($q_{\rm obs}$ and $p_{\rm obs}$, respectively) through uncorrelated Gaussian scatter with unit variance for both observables.

We note that this two-layer Bayesian population model describing the cluster observables is the same as the one that was adopted in a similar SO context in [39].

B. Parameters to be constrained

We assume a flat Λ cold dark matter (ΛCDM) cosmological model and consider a ten-dimensional parameter space. Five of these parameters are cosmological: σ_8, which parametrizes the amplitude of large-scale matter clustering; the physical cold dark matter matter density parameter, $\omega_{\rm c}$; the physical baryon matter density parameter, $\omega_{\rm b}$; the Hubble constant in units of $100\ {\rm km\,s^{-1}\,Mpc^{-1}}$, h; and the spectral index of the primordial scalar perturbations, $n_{\rm s}$. The remaining five parameters are: the tSZ cluster amplitude parameter $A_{\rm SZ}$, the tSZ cluster slope parameter $\alpha_{\rm SZ}$, the scatter in the tSZ signal to noise $\sigma_{\rm SZ}$, the CMB lensing bias parameter $\beta_{\rm CMBlens}$, and the scatter in the CMB lensing signal to noise, $\sigma_{\rm CMBlens}$. From a cosmological point of view, they can be thought of as nuisance parameters to be marginalized over.

C. Synthetic data generation

We generate our synthetic catalogs using the catalog generator of the publicly available COSMOCNC[3] code [39]. At a given point in parameter space, the catalog generator first samples cluster mass-redshift pairs from the halo mass function, producing a realization of the total population of clusters in the Universe down to a lower mass limit $M_{\rm min} = 5 \times 10^{13} M_\odot$. The halo mass function is efficiently evaluated using the COSMOPOWER neural network emulators [56,57] as implemented in CLASS_SZ [58]. The clusters are assigned random locations in the sky, neglecting the effect of cluster clustering; see Sec. V for how this effect could be included. Then, for each cluster, the scaling relations of the mass observables (in this case, $q_{\rm obs}$ and $p_{\rm obs}$) are applied and random scatter is added. This procedure is repeated twice to account for the two layers of scatter in our model, intrinsic and from measurement errors. The catalog is then obtained by imposing the selection criterion, i.e., by selecting all clusters with $q_{\rm obs} > 5$. We refer the reader to Sec. 4.9.1 of [39] for a more detailed description of the catalog generator.

With $N_{\rm cluster} \simeq 16\,000$ clusters at a *Planck* cosmology [39] and three data points per cluster, our catalogs are too high dimensional for current state-of-the-art SBI methods [40]. To address this problem, we compress each catalog by binning the $q_{\rm obs}$–z pairs into a 4×5 grid, defined by five logarithmically spaced bin edges between $q_{\rm obs} = 5$ and $q_{\rm obs} = 40$ and six linearly spaced bin edges between $z = 0.01$ and $z = 1.5$. This two-dimensional grid is chosen to encompass most SO clusters for a *Planck* cosmology (see, e.g., Fig. 2 of [39]). In addition, the CMB lensing signal-to-noise measurements are averaged across the entire catalog, producing a single stacked measurement $\hat{p}_{\rm stacked}$. This choice is justified given the very small per-cluster signal to noise of this observable (see [39]). In brief, each initial $3 \times N_{\rm cluster}$-dimensional data vector is therefore reduced to a 21-dimensional data vector (20 counts in cells and a single stacked mass measurement). Given the comparisons between the binned, unbinned, and unbinned + stacked number count likelihoods presented in [39], this data compression procedure is effectively lossless for SO in a ΛCDM cosmology.

To train our inference method, we generate and compress catalogs for $N_{\rm sim} = 10000$ points in parameter space. These points are drawn from the parameter priors we adopt, which are specified in Table I in Appendix A. In particular, we note that we impose wide uniform priors on the two cosmological parameters that cluster number counts can constrain precisely, σ_8 and $\omega_{\rm c}$. Similarly, we impose wide uniform priors on $A_{\rm SZ}$ and $\alpha_{\rm SZ}$, leaving these parameters to

[2] github.com/simonsobs/so_noise_models/blob/master/LAT_lensing_noise/lensing_v3_1_1/nlkk_v3_1_0_deproj0_SENS1_fsky0p4_qe_lT30-3000_lP30-5000.dat

[3] github.com/inigozubeldia/cosmocnc

be constrained by the CMB lensing data. We impose Gaussian priors on the remaining parameters.

Finally, we generate and compress a set of 100 test catalogs at what we regard as our true point in parameter space. The true parameter values are listed in Table I.

D. Inference methods

We learn the posterior of our compressed catalog data using the neural posterior estimation method of [59], as implemented in the SBI package [60], where it corresponds to the NPE_C method, leaving the exploration of other SBI methods (e.g., [61,62]) to further work. We perform a single training iteration (i.e., obtain an amortized posterior estimate), adopt the MAF density estimator, and terminate the training after 50 epochs of no improvement in the validation loss. We then obtain parameter constraints by drawing 10^5 samples from the learned posterior. We assess the impact of the learning rate η by training with $\eta = 5 \times 10^{-3}$, 5×10^{-4}, 5×10^{-5}, and 5×10^{-6}. We also investigate the impact of the number of training catalogs $N_{\rm sim}$ by training with $N_{\rm sim} = 500$, 1 000, 2 000, 5 000, and 10 000.

In addition, we infer the posterior of one of our test cluster catalogs (our "reference test catalog") following a traditional likelihood-based approach, adopting the same priors as in our SBI analysis. Specifically, we use the COSMOCNC unbinned + stacked likelihood, which provides an almost exact likelihood of our synthetic catalogs [see [39], in particular their Secs. 3.3 and 3.7, for a detailed description of the likelihood; a final expression for the number-count part of the likelihood is provided in their Eq. (18) and for the stacked part in Eq. (29)]. The only approximation is the assumption of Gaussian scatter in the stacked observable, which is highly accurate given the large number of objects being averaged over [39]. This likelihood has been validated in a similar SO context (see [39]). We sample from the posterior using the Markov chain Monte Carlo (MCMC) COBAYA package [63,64].

E. Results

Figure 1 shows the parameter constraints obtained for our reference test catalog following: (i) our SBI approach with our "baseline configuration" ($N_{\rm sim} = 10000$ and $\eta = 5 \times 10^{-4}$; filled blue contours); and (ii) our likelihood-based approach (orange contours). The agreement between the two methods is excellent, with the differences in the mean inferred values $\Delta\mu \equiv \mu_{\rm SBI} - \mu_{\rm like}$ being less than 0.2σ for all parameters (see Table I in Appendix A). Notably, the difference is of only 0.081σ for σ_8. We regard these differences as negligible. We note that the constraints on σ_8, $\omega_{\rm c}$, $A_{\rm SZ}$, and $\alpha_{\rm SZ}$ are data driven, whereas those for the remaining parameters are prior driven and therefore they peak very close to the prior peak. We also note that the observed agreement implies that, for our setup, the choice of bins in the $q_{\rm obs}$–z plane in the SBI-based analysis does not reduce the constraining power of the analysis, as the likelihood is unbinned.

In addition, the two upper panels of Fig. 2 show $\Delta\mu$ in units of the standard deviation of the likelihood-based constraint $\sigma_{\rm like}$ as a function of $N_{\rm sim}$ and η for our two cosmological parameters of interest, σ_8 and $\omega_{\rm c}$ (left and right panel, respectively). The shaded region corresponds to a $\pm 0.2\sigma$ deviation from the likelihood constraint. Defining convergence in our SBI constraints as a change of less than 0.1σ in the constraints on both parameters when doubling $N_{\rm sim}$, this is achieved for $\eta = 5 \times 10^{-3}$ and 5×10^{-4} at $N_{\rm sim} = 10000$. For these configurations, the constraints in both parameters are within 0.2σ of their likelihood-based counterparts. Furthermore, for $N_{\rm sim} > 5000$ and $\eta = 5 \times 10^{-3}$, 5×10^{-4}, and 5×10^{-5}, convergence of the derived constraints within 0.2σ of each other is observed. A learning rate of 5×10^{-6}, however, generally results in poorer performance. We note that due to the stochasticity in the training process (random initialization of the neural network and batch gradient descent), $\Delta\mu/\sigma_{\rm like}$ and $\sigma_{\rm SBI}/\sigma_{\rm like}$ (see below) do not change monotonically with $N_{\rm sim}$.

The lower panels of Fig. 2 show the ratio between the standard deviation of the SBI constraint, $\sigma_{\rm SBI}$, and that of the likelihood-based constraint, $\sigma_{\rm like}$, for the same two cosmological parameters. For σ_8, the standard deviations agree within 5% for $\eta = 5 \times 10^{-3}$, 5×10^{-4}, and 5×10^{-5} when $N_{\rm sim} > 2000$, while for $\omega_{\rm c}$ the agreement is within 10%.

Finally, we perform a self-standing validation of our SBI approach by obtaining SBI parameter constraints for our 100 test catalogs evaluating the posterior obtained with our baseline configuration, which we recall is amortized. Averaging the parameter posterior mean values across all test catalogs, we find biases of less than 0.1σ for all ten parameters (see Table I in Appendix A). In particular, for σ_8 and $\omega_{\rm c}$ no biases are detected at levels of 0.086σ and 0.063σ, corresponding to nondetections at the 0.051% and 0.27% levels, respectively.

F. Goodness of fit

Synthetic data generators naturally enable the evaluation of the goodness of fit of an analysis. This can be achieved by generating a set of data vectors at the "best-fit" point in parameter space. The full data vector distribution at that point can then be mapped and the consistency of the observed data vector with it can be quantified. Alternatively, goodness-of-fit summary statistics can be computed.

Here, we assess the goodness of fit of the SBI-based analysis of our reference SO-like catalog by generating and compressing 100 synthetic catalogs at the "best-fit" point in parameter space (our "best-fit" catalogs) and evaluating

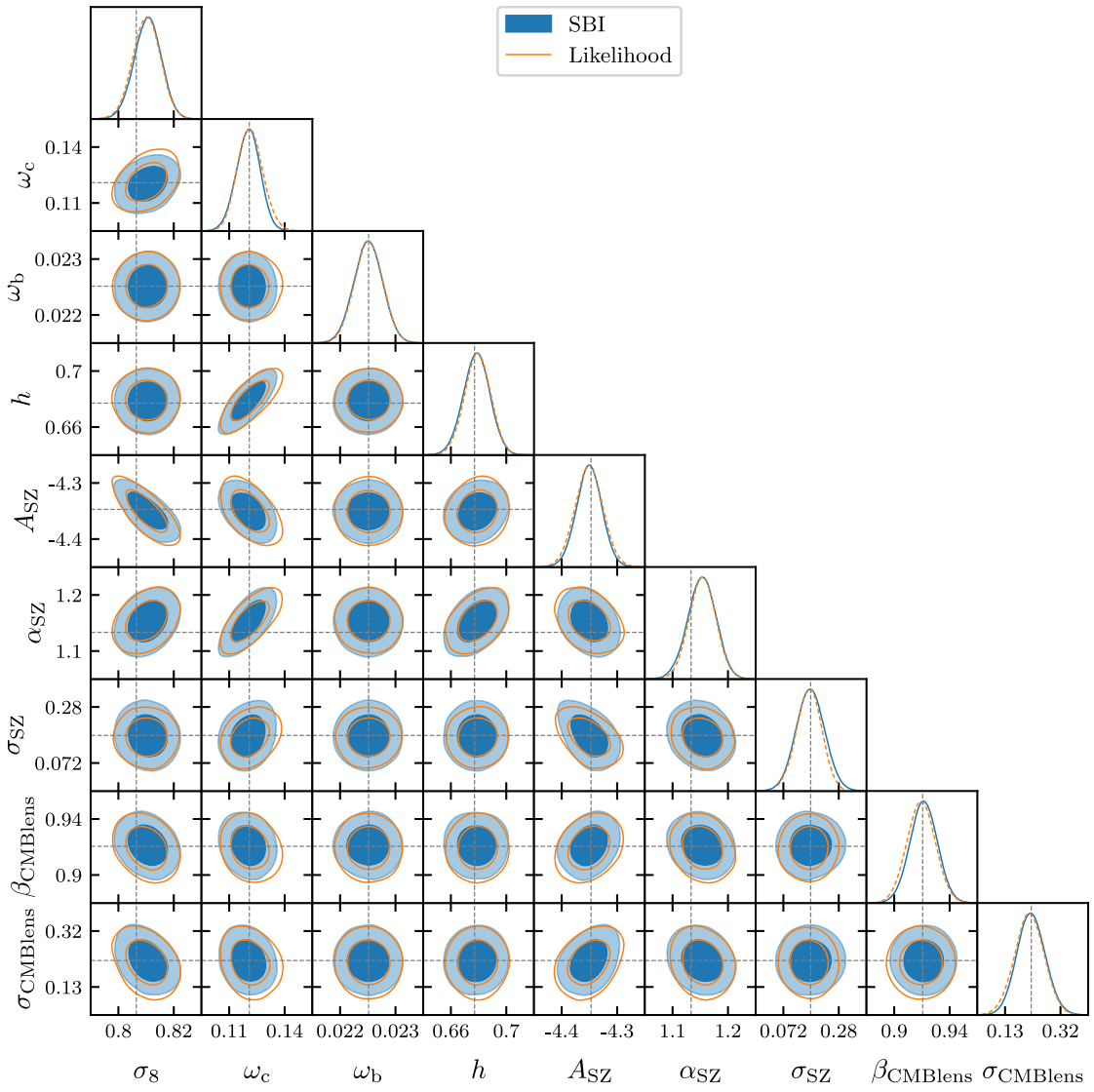

FIG. 1. Parameter constraints obtained with our SBI-based approach (blue filled contours) and our likelihood-based approach (orange contours) for our reference SO-like catalog. The constraints from both methods agree to within 0.2σ for all parameters (see Table I). The dashed lines indicate the true parameter values.

two statistics. These statistics are: (i) the modified Cash statistic of [65], $\hat{C}$, which can be used to assess the goodness of the fit in the number counts by comparing its value to its predicted mean and standard deviation (see, e.g., [14,39]); and (ii) the value of the stacked observable, $\hat{p}_{\rm stacked}$. We take the best-fit point to be specified by the posterior means, noting that in other scenarios in which the posterior is more non-Gaussian, other more suitable points, e.g., the maximum *a posteriori* point, can easily be chosen, as the posterior can be evaluated at any point in parameter space. Alternatively, in a fully Bayesian approach, the best-fit statistics could be evaluated for synthetic catalogs generated at samples from the full posterior distribution [66].

In order to compute $\hat{C}$ and its expected value and standard deviation, a theoretical prediction for the number count in each cell is required. We obtain this prediction by averaging the counts over our 100 best-fit catalogs. Doing this, we obtain an observed value of $\hat{C} = 23.05$ and a predicted mean and standard deviation of $\bar{C}_{\rm sim} = 20.19$ and $\sigma_{C,\rm sim} = 6.35$, respectively. We can therefore conclude that the best-fit prediction is a good fit to the data, as $\hat{C}$ is consistent with its predicted mean $\bar{C}_{\rm sim}$. For comparison, we also evaluate this statistic using a simulation-free prediction for the number counts produced with COSMOCNC, finding $\hat{C} = 21.64$, $\bar{C}_{\rm theory} = 20.18$, and $\sigma_{C,\rm theory} = 6.36$, in good agreement with the simulation-derived values. We note that, in addition to evaluating this statistic, the difference between the (simulation- or theory-based) predicted and observed number counts could also be inspected visually.

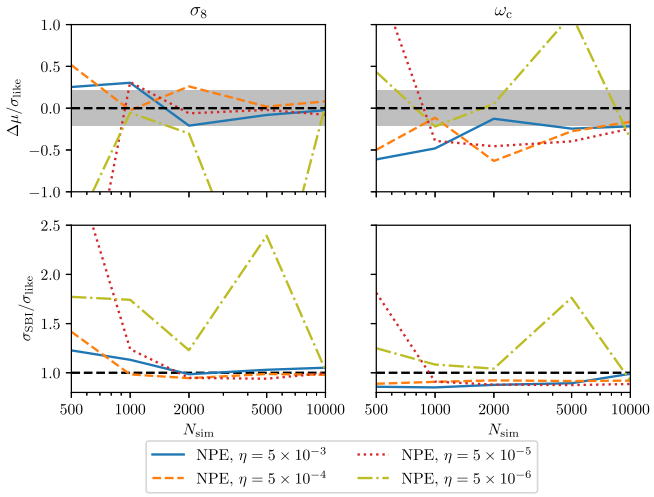

FIG. 2. *Upper panels*: Difference in the mean inferred values between our SBI-based and likelihood-based approaches, $\Delta\mu \equiv \mu_{\rm SBI} - \mu_{\rm like}$, expressed in units of the standard deviation of the likelihood-based constraint $\sigma_{\rm like}$, as a function of $N_{\rm sim}$ and the learning rate η for our two cosmological parameters of interest. The shaded region represents a $\pm 0.2\sigma$ deviation from the likelihood constraint. *Lower panels*: Ratio of the standard deviation of the SBI constraint, $\sigma_{\rm SBI}$, to that of the likelihood constraint, $\sigma_{\rm like}$.

On the other hand, the observed value of the stacked observable is $\hat{p}_{\rm stacked} = 0.3082$. We estimate the mean and standard deviation of $\hat{p}_{\rm stacked}$ by computing its sample mean and standard deviation across the 100 best-fit catalogs. We find $\bar{p}_{\rm stacked,sim} = 0.3105 \pm 0.0009$ (empirical standard deviation) and $\sigma_{p_{\rm stacked},\rm sim} = 0.0085$, in agreement with with the observed value. For comparison, we also compute these two quantities with COSMOCNC (see [39]), finding $\bar{p}_{\rm stacked,theory} = 0.3106$ and $\sigma_{p_{\rm stacked},\rm theory} = 0.0080$, in excellent agreement with the simulation-derived values.

These results demonstrate that the goodness of fit of an SBI-based cluster number count analysis that includes both counts and stacked observables can be quantified using the same synthetic catalog generator used to produce the training catalogs. This implies that our SBI-based approach can be successfully applied to data in a fully self-contained way, without the need to resort to an explicit likelihood at any point in the analysis.

III. APPLICATION TO REAL DATA FROM *PLANCK*

A. The *Planck* cluster catalog and model

We apply our SBI-based approach to the cosmology sample of the *Planck* MMF3 galaxy cluster catalog, which is publicly available on the Planck Legacy Archive[4] (PLA). This sample consists of 439 clusters detected through their tSZ signature using data from the *Planck* experiment [41]. It was used to derive cosmological constraints in the 2015 cluster number count analysis carried out by the *Planck* Collaboration ([13]; hereafter, the "official *Planck* analysis"), and has subsequently been reanalyzed in several studies (e.g., [15,67]).

In the MMF3 cosmology sample, every cluster has a tSZ signal-to-noise measurement, $q_{\rm obs,\,Planck}$, and a redshift measurement, z (the latter except for six clusters). The sample is constructed by imposing a signal-to-noise threshold of 6. This setup is thus very similar to the SO-like one considered in Sec. II.

Here, we adopt the same mass-observable model that was assumed in the official *Planck* analysis, which is also very similar to that of the SO tSZ signal-to-noise adopted in Sec. II. In this model, $q_{\rm obs,\,Planck}$ is linked to the cluster mass and redshift with a two-layer hierarchical model. In the first layer, the mean signal to noise, $\bar{q}_{Planck}(M_{500}, z)$, is given by

$$\bar{q}_{Planck}(M_{500}, z) = \frac{\bar{Y}_{500}((1-b)M_{500}, z)}{\sigma_Y(\theta_{500}((1-b)M_{500}, z))}, \quad (4)$$

where $\bar{Y}_{500}((1-b)M_{500}, z)$ is the cluster's mean integrated Compton-y value, σ_Y is the detection noise (which depends on the sky position and is available on the PLA), and $1-b$ is the tSZ mass bias (also known as the "hydrostatic mass bias"). This scaling relation is equivalent to that of the SO signal to noise, shown in Eq. (1), the difference being that here, following [13], it is parametrized in terms of the integrated Compton-y signal $\bar{Y}_{500}$, instead of the cluster's central Compton-y signal y_0. In order to keep the same notation as in the official *Planck* analysis, we denote the tSZ mass bias with $1-b$ rather than $\beta_{\rm SZ}$ as in Sec. II. The mean integrated Compton-y value, $\bar{Y}_{500}(M_{500}, z)$, is in turn given by

$$\bar{Y}_{500}(M_{500}, z) = Y_\star (h/0.7)^{-2+\alpha} \times \left(\frac{M_{500}}{6 \times 10^{14} M_\odot}\right)^{\alpha} \left(\frac{D_{\rm A}(z)}{10^{-2}\ {\rm Mpc}}\right)^{-2} E^{\beta}(z), \quad (5)$$

where $Y_\star$ parametrizes the tSZ signal amplitude (analogous to $A_{\rm SZ}$ in Sec. II), α its mass dependence (analogous to $\alpha_{\rm SZ}$), and β its redshift dependence, which, following [13], we fix to $\beta = 0.66$.

The logarithm of the mean tSZ signal to noise is then linked to the logarithm of the true tSZ signal-to-noise $\ln q$ through Gaussian intrinsic scatter with a standard deviation of $\sigma_{\ln Y}$ (analogous to $\sigma_{\rm SZ}$ in Sec. II). In the second layer of the model, $\ln q$ is exponentiated and linked to the observed value $q_{\rm obs,Planck}$ through Gaussian scatter with unit variance, which accounts for observational noise.

[4] pla.esac.esa.int

As in Sec. II, we compress the cluster sample by binning the $q_{\mathrm{obs},\,Planck}$–$z$ pairs into a 4×5 grid defined by five logarithmically spaced bin edges between $q_{\mathrm{obs}} = 6$ and $q_{\mathrm{obs}} = 40$ and six linearly spaced bin edges between $z = 0.01$ and $z = 1$, spanning all but one of the clusters in the MMF3 cosmology sample. This procedure compresses the cluster sample into a 20-dimensional data vector. Our grid is coarser than that used in the official *Planck* analysis, but we do not expect this to cause any significant loss of constraining power given the smoothness of the $q_{\mathrm{obs},\,Planck}$–$z$ distribution.

B. Cosmological constraints

As a demonstration of our SBI-based approach on real data, we set out to reproduce the cosmological constraints of the official 2015 *Planck* analysis of [13]. For simplicity, among their four main analyses, we consider their CCCP + H_0 + BBN analysis, as the other three involve combining the cluster catalog with baryon acoustic oscillation data. In this analysis, the parameter space is nine-dimensional and is spanned by: σ_8, Ω_{m} (the matter density parameter), ω_{b}, h, n_{s}, $1 - b$, $\log Y_\star$, α, and $\sigma_{\ln Y}$. Here, we explore the same parameter space and adopt the same priors as in [13], which can be found in Table II of Appendix A. Notably, the weak-lensing-derived prior on the tSZ mass bias is $1 - b = 0.780 \pm 0.092$ (mean and standard deviation; from the Canadian Cluster Comparison Project, CCCP, [68]), and the supernova-derived prior on the Hubble constant is $H_0 = 73.8 \pm 2.4\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ [69]. We emphasize that these priors are adopted solely for the purpose of reproducing the results of [13].

We apply the SBI-based approach described and validated in Sec. II to the compressed MMF3 sample, generating and compressing 5 000 synthetic *Planck* catalogs. We note that these catalogs are generated properly accounting for the sky dependence of the detection noise. We adopt a learning rate of $\eta = 5 \times 10^{-4}$ and stop the training after 50 epochs of no improvement in the validation loss.

We also infer the posterior of the compressed *Planck* data vector using COSMOCNC's binned likelihood, sampling from the posterior with COBAYA. This likelihood is a Poisson likelihood that assumes the same model as the one adopted to generate the catalogs used in the SBI-based approach. Therefore, it is expected to deliver the same posterior as the SBI-based approach. It is also equivalent to the "two-dimensional" likelihood of the official *Planck* analysis (apart from the different binning resolution).

Figure 3 shows the parameter constraints obtained from the MMF3 cosmology sample using our SBI-based approach (solid blue contours) and the likelihood-based approach (orange contours). The agreement between the two methods is excellent, noting that the mildly non-Gaussian shapes of the two-dimensional posteriors are remarkably similar (see also Table II, which lists the parameter constraints for both methods and their differences). For our SBI-based approach, we constrain $\sigma_8 = 0.756 \pm 0.034$, $\Omega_{\mathrm{m}} = 0.342 \pm 0.035$, and $\alpha = 2.05 \pm 0.13$ (mean and standard deviation), noting that the remaining parameters are prior driven (see Table II). These mean values are 0.01, 0.023, and 0.063σ away from their likelihood-based counterparts, which constitutes a remarkable agreement. We also find $\sigma_8(\Omega_{\mathrm{m}}/0.3)^{0.3} = 0.777 \pm 0.034$, with both approaches yielding the same value.

Our constraints on σ_8 and Ω_{m} are consistent within 1σ with those reported in the 2015 *Planck* paper for the same CCCP + H_0 + BBN analysis setup [13], for which $\sigma_8 = 0.78 \pm 0.04$ and $\Omega_{\mathrm{m}} = 0.31 \pm 0.04$. There is better agreement for $\sigma_8(\Omega_{\mathrm{m}}/0.3)^{0.3}$, for which [13] report $\sigma_8(\Omega_{\mathrm{m}}/0.3)^{0.3} = 0.772 \pm 0.034$, 0.15σ away from our constraint. Given that the constraints in [13] are obtained for the same data, model, and priors as in our analysis, the observed differences are likely due to differences in implementation. Since the goal of this exercise is to demonstrate that our SBI-based approach can be applied to real data (and yields constraints that are virtually identical to those obtained in our own likelihood-based analysis), we do not investigate these differences further here.

IV. COMPARISON TO PREVIOUS WORK

The use of SBI in galaxy cluster cosmology was first proposed, as a proof of concept, in [70], in which the approximate Bayesian computation (ABC) method was applied. More recently, [71] applied both ABC and the PYDELFI algorithm [72] to optically selected synthetic cluster catalogs, assessing their performance to the 1σ level with one test catalog. In addition, [73] applied the sequential neural posterior estimator (SNPE) of [74] to X-ray-selected synthetic catalogs, finding no evidence for biases in cosmological parameters as the constraining power of the test catalog was increased. Finally, [75] applied the same SNPE method to optically selected synthetic catalogs, finding no biases in the relevant cosmological parameters to the 1σ level and consistency with a likelihood-based approach similarly to the $0.5 - 1\sigma$ level. In addition, they also applied, for the first time, an SBI approach to synthetic cluster catalogs produced using halo catalogs from a cosmological simulation (specifically, from the QUIJOTE simulations; [76]).

The current paper takes the work presented in [70,71,73,75] further in several ways. First, we constrain the biases of all the parameters to less 0.1σ in the highly demanding context of SO, improving upon what has been generally demonstrated previously (with the exception of [73], where no biases were found when the constraining power of the test catalog was increased by a very large amount, noting that the comparison is difficult given that [73] do not quantify the parameter biases as we do). We also find agreement between our SBI and likelihood-based

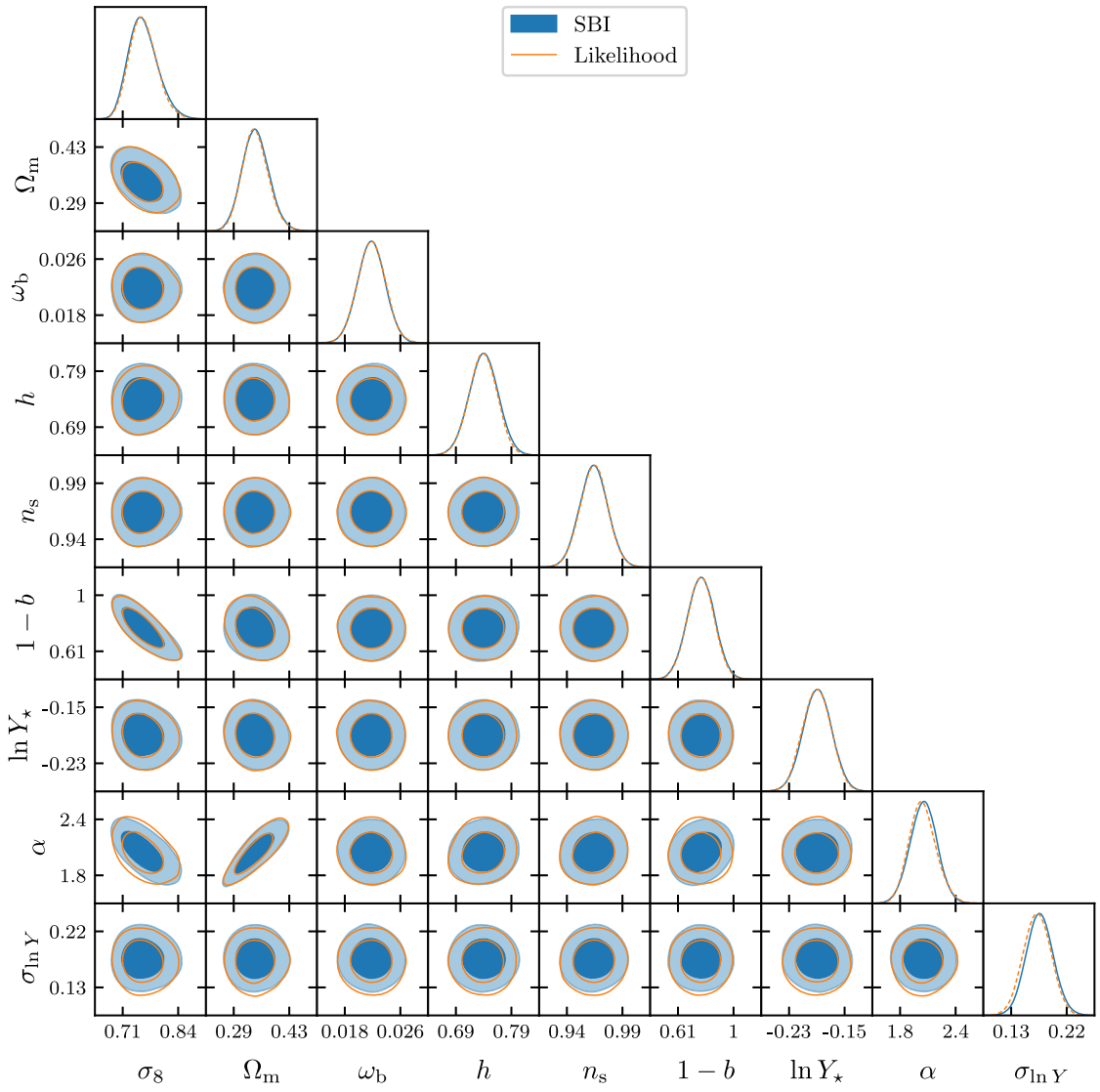

FIG. 3. Parameter constraints for the *Planck* MMF3 cosmology sample (real data), obtained with both our SBI-based approach (filled blue contours) and our likelihood-based approach (orange contours). The agreement between both methods is excellent, with differences smaller than 0.07σ for the posterior means of all parameters (see Table II) and with the mildly non-Gaussian shapes of the two-dimensional posteriors being remarkably similar. We emphasize that the adopted priors (including those on h and $1-b$) were chosen in order to have the same setup as in one of the four main analyses in the *Planck* Collaboration 2015 cluster number count paper [13].

constraints to less than 0.17σ for SO and less than 0.07σ for real *Planck* data, improving upon what is reported in [75] (the only other analysis carrying out such a comparison) for their synthetic halo-mass-function-based catalogs (agreement at the $0.5-1\sigma$ level). In contrast, unlike in [75], here we do not attempt to use halo catalogs from cosmological simulations, leaving the exploration of this more realistic avenue to further work. We also apply a modern neural-estimation-based method to a tSZ catalog for the first time. In addition, we propose and validate, for the first time, an approach to quantify the goodness of fit of an SBI-based cluster number count analysis that relies solely on synthetic catalogs; goodness of fit is not discussed in [70,71,73,75]. Finally, and importantly, the application of our SBI-based approach to the *Planck* MMF3 cosmology sample constitutes, to the authors' knowledge, the first application of SBI to carry out a number count analysis of a real galaxy cluster catalog.

V. THE CASE FOR SBI-BASED GALAXY CLUSTER COSMOLOGY

The results presented in Sec. II demonstrate that our SBI-based approach for galaxy cluster number count cosmological inference is able to deliver unbiased cosmological constraints in the highly demanding context of SO and provides a natural way to quantify the goodness of fit. In addition, in Sec. III we have demonstrated that it can be

successfully applied to real data. All of this is achieved without the need to develop and implement an explicit likelihood, relying solely on synthetic cluster catalogs.

In the two specific setups we considered we already had an explicit likelihood implementation. However, we argue that even when a likelihood can be formulated and implemented, our SBI-based approach offers several significant advantages, as follows.

(1) Cluster number count likelihoods are usually very complex, often involving a large number of multidimensional integrals to marginalize over the latent variables in the Bayesian hierarchical population models describing the cluster catalogs (see, e.g., [12,14,15,28,38,39]). Developing and validating such likelihoods requires substantial investment, often taking months to years. In contrast, generating synthetic catalogs, as done in this work, is significantly easier, requiring only efficient sampling from the halo mass function and random number generation to account for the scatter in the cluster observables.

(2) The complexity of the model describing the cluster observables can significantly impact how efficiently a likelihood can be computed. Challenges include, e.g., correlated scatter between the mass observables and mass-dependent scatter (see, e.g., [39]). Synthetic catalog generation remains unaffected by such complexities[5]: once the scaling relations and the scatter parameters are specified, the same generator can be used for any catalog. Moreover, the generation time is not affected by correlated scatter and the number of mass observables (see Appendix B).

(3) Evaluating cluster number count likelihoods can be computationally expensive. For instance, the COSMOCNC likelihood used in this work takes approximately 12 s per evaluation for our SO-like catalog with two mass observables, neglecting correlated scatter. More mass observables and larger catalogs, both typical scenarios in upcoming experiments, will only exacerbate this problem. Furthermore, traditional MCMC methods offer limited parallelization. In contrast, while generating synthetic catalogs can also be computationally demanding (particularly for accurate sampling of the halo mass function), the process scales minimally with the number of observables (see Appendix B) and is trivially parallelizable *en masse*, as the catalogs are generated independently from each other. As shown in Appendix B, for our SO setup, the SBI analysis achieves a comparable total central processing unit (CPU) time but significantly lower wall-clock time than the likelihood-based approach.

An SBI-based approach along the lines of the one proposed here becomes even more advantageous when an accurate likelihood is either computationally prohibitive or simply intractable. In cluster cosmology, this situation can arise when sample variance due to cluster clustering cannot be neglected, making the Poissonian likelihood used here inaccurate (see, e.g., [77]). While accounting for sample variance in a binned likelihood with a single mass observable is relatively straightforward (see, e.g., [7]), doing so in more complicated setups, e.g., when additional mass observables are used for mass calibration (as in this work for the SO-like survey), is significantly more challenging [78]. However, incorporating the effect of sample variance into synthetic cluster catalogs is a simpler problem. This could be achieved, e.g., by sampling from a density-field-modulated halo mass function or by directly using cluster catalogs from cosmological simulations (e.g., [76,79]). These cluster catalogs could then be used to generate synthetic catalogs by applying the cluster observable model to them (see, e.g., [75]). Further realism could be achieved by using catalogs produced by applying the cluster finder used to obtain the real catalog to simulated sky maps. Doing this would account for sample variance due to cluster clustering in the inference process, as well as for other complications such as blending of several clusters into a single detection. It would also allow one to drop simplifying assumptions in the model, such as log-normal intrinsic scatter and power-law scaling relations. However, if cosmological simulations different from or in addition to those available in existing suites (e.g., [76]) are needed, this realism would likely be achieved at the expense of a significantly higher computational cost for training and a potentially more challenging training process. We leave the exploration of these promising approaches to future work.

Moreover, SBI naturally facilitates the combination of cluster catalogs with correlated datasets, such as cosmic shear or tSZ power spectrum data, provided that consistent simulations exist. Computing joint likelihoods of correlated datasets, however, can be a difficult task (e.g., [80]).

Finally, we remark that our SBI-based approach can be validated, under the assumption of the adopted model, using the same catalog generator employed to produce the training data, as demonstrated in Sec. II E. Similar catalogs can be used to quantify the goodness of fit of an analysis, as shown in Sec. II F. This means that a cosmological analysis can be fully carried out without the need to resort to explicit likelihoods at any point. We note that cluster number count likelihoods are often validated with synthetic catalogs as well (e.g., [15,38,39]), which means that a catalog generator is often already part of a typical likelihood-based cluster cosmology inference pipeline.

[5]With the exception of non-Gaussian scatter (not including log-normal scatter, as it is Gaussian in the logarithm of the variable), for which a tailored random number generator would have to be used.

Given the significant advantages outlined here, we believe that SBI has the potential to be transformative for galaxy cluster cosmology as a field. By significantly reducing both development and computational costs, it offers a highly promising pathway for obtaining cosmological constraints from galaxy cluster catalogs from current and upcoming experiments.

VI. CONCLUSION

In this work we have implemented an SBI-based approach for galaxy cluster number count cosmological inference. We have validated it in the context of the Simons Observatory, showing that it delivers unbiased cosmological constraints within the ΛCDM model, with any parameter biases constrained to below 0.1σ. We have then applied it to the real *Planck* MMF3 cosmology sample, finding cosmological constraints that are in excellent agreement—better than 0.07σ—with their likelihood-based counterparts from our own reanalysis. We expect this approach to be directly applicable to other real datasets, regardless of the nature of the mass observables, i.e., whether obtained through X-ray, optical or mm observations, and whether they are scalar or vector quantities (see [70,71,73,75]). This includes catalogs from other current and upcoming experiments such as *eROSITA*, *Euclid*, Rubin/LSST, SPT-3G, and CMB-S4. We also expect our approach to be directly applicable for constraining any extensions to the ΛCDM model to which cluster number counts are sensitive. In particular, it could be used to set constraints on the dark energy equation of state and the sum of the neutrino masses.

As discussed in detail in Sec. V, in galaxy cluster cosmology, an SBI-based approach like ours presents several major advantages over the traditional explicit likelihood-based approach: (i) it requires significantly less development effort; (ii) it can accommodate complicated models more easily; (iii) it allows for much greater computational parallelization; and (iv) it has the potential to incorporate sample variance due to cluster clustering in a much more straightforward way. In addition, a given analysis setup can be easily validated with the generation of a set of test catalogs, as demonstrated in Sec. II E, and the goodness of fit can be quantified using a similar set of catalogs, as shown in Sec. II F. Thus, as we have demonstrated for the first time, a cluster number count cosmological analysis can be fully carried out without resorting to an explicit likelihood at any stage. Given these advantages, we believe that SBI has the potential to be transformative for galaxy cluster cosmology as a field, significantly simplifying and accelerating the analysis of galaxy cluster catalogs from current and upcoming experiments.

This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service,[6] provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant no. EP/T022159/1), and DiRAC funding from the Science and Technology Facilities Council,[7] within the DiRAC Cosmos dp002 project. Figures 1 and 3 were produced using the GETDIST package [81].

ACKNOWLEDGMENTS

The authors would like to thank Boryana Hadzhiyska, Licong Xu, Peter Melchior, Joseph J. Mohr, Sebastian Bocquet, Matthias Klein, James Alvey, Aditya Singh, and Sophie Vogt for useful discussions, and Jean-Baptiste Melin for comments on the manuscript. Í. Z. and A. C. acknowledge support from the STFC (grants no. ST/W000977/1 and no. ST/X006387/1).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: PARAMETER PRIORS, CONSTRAINTS AND BIASES

Table I lists the true parameter values (second column), adopted priors (third column), standard deviations of the likelihood-based analysis $\sigma_{\rm like}$ (fourth column), differences between the SBI-based and likelihood-based constraints in units of $\sigma_{\rm like}$ for our reference SO-like test catalog, $(\mu_{\rm SBI} - \mu_{\rm like})/\sigma_{\rm like}$ (fifth column), and constraints on the parameter biases, both in units of the SBI standard deviation $\sigma_{\rm SBI}$ and in percentage (sixth and seventh columns, respectively), for the analysis demonstration with our SO-like synthetic catalogs. Notably, $(\mu_{\rm SBI} - \mu_{\rm like})$ is less than $0.2\sigma_{\rm like}$ for all parameters, the largest difference being at $-0.163\sigma_{\rm like}$ for $\omega_{\rm c}$. We regard these differences as negligible. Furthermore, all the biases are less than 0.1σ.

We note that all the parameters on which a Gaussian prior is imposed are prior driven, their likelihood-based standard deviations being equal to those of the corresponding priors. As a consequence, the statistical errors on their biases are very small, as the peak of the posterior is located very close to the peak of the prior in all realizations. The remaining four parameters, on the other hand, are data driven, with posterior standard deviations significantly smaller than those of the prior and with relatively looser constraints on their biases. These bias constraints, however, are still at the 0.1σ level or better.

In addition, Table II shows the priors, the SBI-based and likelihood-based posteriors (mean and standard deviation), and the differences between the posteriors for the real

[6](csd3.cam.ac.uk)
[7](dirac.ac.uk)

TABLE I. True parameter values, priors, standard deviation of the likelihood-based analysis $\sigma_{\rm like}$, difference between the SBI-based and likelihood-based posterior means in units of $\sigma_{\rm like}$ for our reference test catalog, $(\mu_{\rm SBI}-\mu_{\rm like})/\sigma_{\rm like}$, and constraints on the parameter biases, both in units of the SBI standard deviation $\sigma_{\rm SBI}$ and in percentage, for the analysis demonstration with our SO-like synthetic catalogs (see also Fig. 1). For all the parameters, $(\mu_{\rm SBI}-\mu_{\rm like})/\sigma_{\rm like}$ is less than 0.2 and the parameter biases are constrained to being less than 0.1σ. Note that constraints on the first four parameters are data driven, while those on the remaining six parameters are prior driven.

Parameter	True value ($p_{\rm true}$)	Prior	$\sigma_{\rm like}$	$(\mu_{\rm SBI}-\mu_{\rm like})/\sigma_{\rm like}$	$\langle(\mu_{\rm SBI}-p_{\rm true})\rangle/\langle\sigma_{\rm SBI}\rangle$	$\langle(\mu_{\rm SBI}-p_{\rm true})/p_{\rm true}\rangle$ [%]
σ_8	0.811	$\mathcal{U}(0.7,0.9)$	0.0051	0.081	-0.107 ± 0.086	-0.064 ± 0.051
$\omega_{\rm c}$	0.121	$\mathcal{U}(0.0804,0.204)$	0.0060	−0.163	0.031 ± 0.063	0.13 ± 0.27
$A_{\rm SZ}$	−4.31	$\mathcal{U}(-4.41,-4.21)$	0.028	0.0102	0.052 ± 0.076	-0.030 ± -0.043
$\alpha_{\rm SZ}$	1.12	$\mathcal{U}(1,1.24)$	0.016	−0.0593	0.10 ± 0.10	0.13 ± 0.15
h	0.674	$\mathcal{N}(0.674,0.01)$	0.010	−0.0867	0.0211 ± 0.0033	0.0309 ± 0.0048
$\omega_{\rm b}$	0.0222	$\mathcal{N}(0.0222,0.00015)$	0.00015	−0.0373	-0.1039 ± 0.0023	-0.0689 ± 0.0015
$n_{\rm s}$	0.96	$\mathcal{N}(0.96,0.0042)$	0.0042	−0.0603	-0.0395 ± 0.0015	-0.01767 ± 0.00068
$\sigma_{\rm SZ}$	0.173	$\mathcal{N}(0.173,0.05)$	0.045	0.101	-0.0585 ± 0.0063	-1.64 ± 0.18
$\beta_{\rm CMBlens}$	0.92	$\mathcal{N}(0.92,0.01)$	0.010	0.182	-0.0122 ± 0.0048	-0.0133 ± 0.0052
$\sigma_{\rm CMBlens}$	0.22	$\mathcal{N}(0.22,0.05)$	0.050	0.127	-0.0417 ± 0.0048	-0.90 ± 0.10

TABLE II. Priors, SBI-based and likelihood-based posteriors (mean and standard deviation), and differences between the posterior means for the *Planck* MMF3 cosmology sample for the CCCP + H_0 + BBN prior combination. The agreement between the two sets of parameter constraints is excellent, with differences of less than 0.07σ for all parameters. We emphasize that the adopted priors are chosen specifically to reproduce the official 2015 *Planck* analysis of [13].

Parameter	Prior	SBI posterior mean $\mu_{\rm SBI}$	Likelihood posterior mean $\mu_{\rm like}$	$(\mu_{\rm SBI}-\mu_{\rm like})/\sigma_{\rm like}$
σ_8	$\mathcal{U}(0.6,0.9)$	0.756 ± 0.036	0.756 ± 0.034	0.010
$\Omega_{\rm m}$	$\mathcal{U}(0.2,0.5)$	0.342 ± 0.035	0.341 ± 0.034	0.023
α	$\mathcal{U}(1,3)$	2.05 ± 0.13	2.04 ± 0.13	0.063
$\omega_{\rm b}$	$\mathcal{N}(0.022,0.002)$	0.0220 ± 0.0020	0.0220 ± 0.0020	−0.0078
h	$\mathcal{N}(0.738,0.024)$	0.740 ± 0.025	0.738 ± 0.024	0.054
$n_{\rm s}$	$\mathcal{N}(0.962,0.014)$	0.962 ± 0.014	0.963 ± 0.014	−0.015
$1-b$	$\mathcal{N}(0.78,0.092)$	0.771 ± 0.096	0.769 ± 0.093	0.020
$\ln Y_\star$	$\mathcal{N}(-0.19,0.02)$	-0.190 ± 0.020	-0.190 ± 0.020	0.048
$\sigma_{\ln Y}$	$\mathcal{N}(0.173,0.023)$	0.177 ± 0.022	0.172 ± 0.023	0.20

Planck MMF3 cosmology sample for the CCCP + H_0 + BBN prior combination. There is excellent agreement between the two sets of parameter constraints, with differences of less than 0.07σ for all parameters. We stress that the adopted priors are chosen specifically to reproduce the official 2015 *Planck* analysis of [13].

We note that σ_8, $\Omega_{\rm m}$, and α are data driven, with the remaining six parameters being prior driven.

APPENDIX B: COMPUTATIONAL COSTS: SBI VS LIKELIHOOD

We first report the total CPU and wall-clock times required to perform the analyses for our SO-like setup. At about 12 s per likelihood evaluation, it took 7.5 days on one node with 76 CPUs for four MCMC chains (19 CPU per chain) to produce a total of 82 030 accepted samples, at which we stopped. This constitutes a total wall-clock time of 180 h and a total CPU time of 13 680 h for the likelihood-based analysis. On the other hand, we used eight nodes, each with 76 CPUs, to generate our 10 000 training catalogs for the SBI analysis. Due to memory constraints, only one in eight CPUs were used in practice. At an average of 441 s per catalog (obtained by averaging over 10 randomly chosen catalogs), this leads to a total wall-clock time of 16.1 h and a total CPU time of 9 807 h. We note that the time taken to train the posterior estimate adds a negligible amount of time to this total (around 1–30 min on one CPU, depending on the number of simulations and learning rate).

That is, the total CPU times are comparable for both methods, with the SBI-based approach being 1.4 times cheaper than its likelihood-based counterpart. However, in terms of wall-clock time, the SBI-based approach is faster by a factor of 11.2, which constitutes an impressive mark.

We note that we have used more computational resources for our SBI-based approach, with eight nodes instead of the single node used in the likelihood-based approach. This choice is motivated by the fact that, for our SO-like setup and computer (CSD3 at the University of Cambridge), the likelihood evaluation time reaches a minimum when using about 15–20 CPUs, and increases beyond that, likely due to communication overheads. The likelihood code is parallelized using Python's MULTIPROCESSING module (see [39] and the code repository). Alternative parallelization schemes such as MPI may reduce communication overheads and enable faster likelihood evaluations if more CPUs are used. In addition, MULTIPROCESSING only allows parallelization within a single node, so we are limited to one node per MCMC chain, though this constraint is not reached in our case.

Given these considerations, we argue that the SBI-based approach allows for easier and more scalable parallelization. With enough CPUs, the wall-clock time for the SBI-based analysis could be reduced to 441 s (in addition to the few minutes required to train the posterior estimate). This parallelization is trivial, as all catalogs are generated independently. In contrast, MCMC methods are inherently sequential. For the MCMC setup we consider (four chains using the Metropolis-Hastings algorithm, as implemented in COBAYA), assuming approximately 100 000 likelihood evaluations are needed for convergence and that extensive parallelization is available via MPI, matching a 441 s wall-clock time would require each likelihood evaluation to take about $441\ \mathrm{s}/100000 \times 4 = 0.018$ s. Given that the COSMOPOWER emulator, which we use to compute the matter power spectrum (used, in turn, to compute the halo mass function), takes about 0.015 s per redshift, achieving this target seems difficult, especially when accounting for communication overheads in a highly parallelized setup. Even if achievable, e.g., by switching to a different MCMC sampler or programming language, it would likely require significant development effort.

Finally, we note that the time required to generate synthetic catalogs is entirely dominated by the process of drawing mass-redshift pairs from the halo mass function. As a result, it scales very mildly with the complexity of the cluster-observable model or the number of mass observables. We have tested two extensions to our SO-like setup: (i) adding correlated intrinsic scatter between the two mass observables $q_{\rm obs}$ and $p_{\rm obs}$; and (ii) introducing eight additional mass observables, all with correlated intrinsic scatter, for a total of ten mass observables. We simply take the additional observables to be copies of $q_{\rm obs}$ and note that all relevant calculations are repeated for each one. Neither extension led to a statistically significant increase in catalog generation time. This stands in sharp contrast with a likelihood-based approach, where adding correlated scatter and more observables typically results in a substantial increase in evaluation time: evaluating the likelihood for around 16000 clusters with ten correlated mass observables would indeed be extremely challenging.

[1] V. R. Eke, S. Cole, and C. S. Frenk, Mon. Not. R. Astron. Soc. **282**, 263 (1996).
[2] R. A. Battye and J. Weller, Phys. Rev. D **68**, 083506 (2003).
[3] S. W. Allen, A. E. Evrard, and A. B. Mantz, Annu. Rev. Astron. Astrophys. **49**, 409 (2011).
[4] D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo, Phys. Rep. **530**, 87 (2013).
[5] N. A. Bahcall and X. Fan, Astrophys. J. **504**, 1 (1998).
[6] S. Borgani, P. Rosati, P. Tozzi, S. A. Stanford, P. R. Eisenhardt, C. Lidman, B. Holden, R. Della Ceca, C. Norman, and G. Squires, Astrophys. J. **561**, 13 (2001).
[7] E. Rozo, R. H. Wechsler, E. S. Rykoff *et al.*, Astrophys. J. **708**, 645 (2010).
[8] M. Hasselfield, M. Hilton, T. A. Marriage *et al.*, J. Cosmol. Astropart. Phys. 07 (2013) 008.
[9] Planck 2013 results XX, Astron. Astrophys. **571**, A20 (2014).
[10] A. B. Mantz, A. von der Linden *et al.*, Mon. Not. R. Astron. Soc. **446**, 2205 (2015).
[11] S. Bocquet, A. Saro, J. J. Mohr *et al.*, Astrophys. J. **799**, 214 (2015).
[12] T. de Haan, B. A. Benson, L. E. Bleem *et al.*, Astrophys. J. **832**, 95 (2016).
[13] Planck 2015 results XXIV, Astron. Astrophys. **594**, A24 (2016).
[14] S. Bocquet, J. P. Dietrich, T. Schrabback, L. E. Bleem, M. Klein *et al.*, Astrophys. J. **878**, 55 (2019).
[15] Í. Zubeldia and A. Challinor, Mon. Not. R. Astron. Soc. **489**, 401 (2019).
[16] B. Bolliet, T. Brinckmann, J. Chluba, and J. Lesgourgues, Mon. Not. R. Astron. Soc. **497**, 1332 (2020).
[17] M. Costanzi, E. Rozo, M. Simet, Zhang *et al.*, Mon. Not. R. Astron. Soc. **488**, 4779 (2019).
[18] M. H. Abdullah, A. Klypin, and G. Wilson, Astrophys. J. **901**, 90 (2020).
[19] T. M. C. Abbott, M. Aguena, A. Alarcon, Allam *et al.*, Phys. Rev. D **100**, 023509 (2020).
[20] C.-H. To, E. Krause, E. Rozo, H.-Y. Wu, D. Gruen, J. DeRose, E. Rykoff, R. H. Wechsler, M. Becker, M. Costanzi, T. Eifler, M. E. da Silva Pereira, and N. Kokron (DES Collaboration), Mon. Not. R. Astron. Soc. **502**, 4093 (2021).

[21] C. Garrel, M. Pierre, Valageas *et al.*, Astron. Astrophys. **663**, A3 (2022).
[22] P. S. Chaubal, C. L. Reichardt, N. Gupta *et al.*, Astrophys. J. **931**, 139 (2022).
[23] G. F. Lesci, F. Marulli, L. Moscardini, Sereno *et al.*, Astron. Astrophys. **659**, A88 (2022).
[24] I. N. Chiu, M. Klein, J. Mohr, and S. Bocquet, Mon. Not. R. Astron. Soc. **522**, 1601 (2023).
[25] T. Sunayama, H. Miyatake, S. Sugiyama *et al.*, Phys. Rev. D **110**, 083511 (2024).
[26] A. Fumagalli, M. Costanzi, A. Saro, T. Castro, and S. Borgani, Astron. Astrophys. **682**, A148 (2024).
[27] S. Bocquet, S. Grandis, L. E. Bleem, M. Klein, J. J. Mohr, Schrabback *et al.*, Phys. Rev. D **110**, 083510 (2024).
[28] V. Ghirardini, E. Bulbul, E. Artis, Clerc *et al.*, Astron. Astrophys. **689**, A298 (2024).
[29] E. Lee, R. Battye, and B. Bolliet, Mon. Not. R. Astron. Soc. **538**, 2768 (2025).
[30] Í. Zubeldia, J.-B. Melin, J. Chluba, and R. Battye, Mon. Not. R. Astron. Soc. **539**, 185 (2025).
[31] DES Collaboration, arXiv:2503.13632.
[32] A. Merloni, P. Predehl, W. Becker *et al.*, arXiv:1209.3114.
[33] R. Laureijs, J. Amiaux, S. Arduini *et al.*, arXiv:1110.3193.
[34] LSST Dark Energy Science Collaboration, arXiv:1211.0310.
[35] B. A. Benson, P. A. R. Ade, Z. Ahmed *et al.*, SPIE Int. Soc. Opt. Eng. **9153**, 91531P (2014).
[36] Simons Observatory Collaboration, J. Cosmol. Astropart. Phys. 02 (2019) 056.
[37] K. N. Abazajian, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, K. S. Arnold, C. Baccigalupi, J. G. Bartlett, N. Battaglia, B. A. Benson *et al.*, arXiv:1610.02743.
[38] S. Bocquet, S. Grandis, L. E. Bleem, M. Klein, J. J. Mohr *et al.*, Phys. Rev. D **110**, 083509 (2024).
[39] Í. Zubeldia and B. Bolliet, J. Cosmol. Astropart. Phys. 11 (2024) 018.
[40] K. Cranmer, J. Brehmer, and G. Louppe, Proc. Natl. Acad. Sci. U.S.A. **117**, 30055 (2020).
[41] Planck 2015 results XXVI, Astron. Astrophys. **594**, A26 (2016).
[42] The Simons Observatory Collaboration, arXiv:2503.00636.
[43] M. Arnaud, G. W. Pratt, R. Piffaretti, H. Böhringer, J. H. Croston, and E. Pointecouteau, Astron. Astrophys. **517**, A92 (2010).
[44] M. Hilton, M. Hasselfield, C. Sifón, N. Battaglia *et al.*, Astrophys. J. Suppl. Ser. **235**, 20 (2018).
[45] Í. Zubeldia, A. Rotti, J. Chluba, and R. Battye, Mon. Not. R. Astron. Soc. **522**, 4766 (2023).
[46] Í. Zubeldia, J. Chluba, and R. Battye, Mon. Not. R. Astron. Soc. **522**, 5123 (2023).
[47] G. Stein, M. A. Alvarez, and J. R. Bond, Mon. Not. R. Astron. Soc. **483**, 2236 (2019).
[48] G. Stein, M. A. Alvarez, J. R. Bond, A. van Engelen, and N. Battaglia, J. Cosmol. Astropart. Phys. 10 (2020) 012.
[49] V. Biffi, S. Borgani, G. Murante, E. Rasia, S. Planelles, G. L. Granato, C. Ragone-Figueroa, A. M. Beck, M. Gaspari, and K. Dolag, Astrophys. J. **827**, 112 (2016).
[50] D. J. Barnes, M. Vogelsberger, F. A. Pearce, A.-R. Pop, R. Kannan, K. Cao, S. T. Kay, and L. Hernquist, Mon. Not. R. Astron. Soc. **506**, 2533 (2021).
[51] F. Jennings and R. Davé, Mon. Not. R. Astron. Soc. **526**, 1367 (2023).
[52] Í. Zubeldia and A. Challinor, Mon. Not. R. Astron. Soc. **497**, 5326 (2020).
[53] J.-B. Melin and J. G. Bartlett, Astron. Astrophys. **578**, A21 (2015).
[54] A. Huchet and J.-B. Melin, Astron. Astrophys. **689**, A303 (2024).
[55] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **490**, 493 (1997).
[56] A. Spurio Mancini, D. Piras, J. Alsing, B. Joachimi, and M. P. Hobson, Mon. Not. R. Astron. Soc. **511**, 1771 (2022).
[57] B. Bolliet, A. Spurio Mancini, J. C. Hill, M. Madhavacheril, H. T. Jense, E. Calabrese, and J. Dunkley, Mon. Not. R. Astron. Soc. **531**, 1351 (2024).
[58] B. Bolliet, A. Kusiak, F. McCarthy *et al.*, EPJ Web Conf. **293**, 00008 (2024).
[59] D. S. Greenberg, M. Nonnenmacher, and J. H. Macke, arXiv:1905.07488.
[60] A. Tejero-Cantero, J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. J. Gonçalves, D. S. Greenberg, and J. H. Macke, J. Open Source Softwaare **5**, 2505 (2020).
[61] G. Papamakarios, D. Sterratt, and I. Murray, in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 89, edited by K. Chaudhuri and M. Sugiyama (PMLR, Cambridge, MA, 2019), pp. 837–848.
[62] J. Alvey, U. Bhardwaj, V. Domcke, M. Pieroni, and C. Weniger, Phys. Rev. D **109**, 083008 (2024).
[63] J. Torrado and A. Lewis, Cobaya: Bayesian analysis in cosmology, Astrophysics Source Code Library, record ascl:1910.019 (2019), ascl:1910.019.
[64] J. Torrado and A. Lewis, J. Cosmol. Astropart. Phys. 05 (2021) 057.
[65] J. S. Kaastra, Astron. Astrophys. **605**, A51 (2017).
[66] M. von Wietersheim-Kramsta, K. Lin, N. Tessore, B. Joachimi, A. Loureiro, R. Reischke, and A. H. Wright, Astron. Astrophys. **694**, A223 (2025).
[67] L. Salvati, A. Saro, S. Bocquet *et al.*, Astrophys. J. **934**, 129 (2022).
[68] H. Hoekstra, R. Herbonnet, A. Muzzin, A. Babul, A. Mahdavi, M. Viola, and M. Cacciato, Mon. Not. R. Astron. Soc. **449**, 685 (2015).
[69] A. G. Riess, L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha, W. Li, and R. Chornock, Astrophys. J. **730**, 119 (2011).
[70] E. E. O. Ishida, S. D. P. Vitenti, M. Penna-Lima *et al.*, Astron. Comput. **13**, 1 (2015).
[71] S.-I. Tam, K. Umetsu, and A. Amara, Astrophys. J. **925**, 145 (2022).
[72] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, Mon. Not. R. Astron. Soc. **488**, 4440 (2019).
[73] M. Kosiba, N. Cerardi, M. Pierre, F. Lanusse, C. Garrel, N. Werner, and M. Shalak, Astron. Astrophys. **693**, A46 (2025).
[74] G. Papamakarios and I. Murray, in *Advances in Neural Information Processing Systems*, edited by D. Lee,

M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., La Jolla, CA, 2016), Vol. 29.
[75] M. Reza, Y. Zhang, C. Avestruz, L. E. Strigari, S. Shevchuk, and F. Villaescusa-Navarro, arXiv:2409.20507.
[76] F. Villaescusa-Navarro, C. Hahn, E. Massara *et al.*, Astrophys. J. Suppl. Ser. **250**, 2 (2020).
[77] W. Hu and A. V. Kravtsov, Astrophys. J. **584**, 702 (2003).
[78] C. Payerne, C. Murray, C. Combet, C. Doux, A. Fumagalli, and M. Penna-Lima, Mon. Not. R. Astron. Soc. **520**, 6223 (2023).
[79] A. E. Bayer, Y. Zhong, Z. Li, J. DeRose, Y. Feng, and J. Liu, J. Cosmol. Astropart. Phys. 05 (2025) 016.
[80] A. Nicola, J. Dunkley, and D. N. Spergel, Phys. Rev. D **102**, 083505 (2020).
[81] A. Lewis, J. Cosmol. Astropart. Phys. 08 (2025) 025.