Sample variance in weak lensing: how many simulations are required?

Andrea Petri,1,2 Zoltán Haiman,3 and Morgan May2

1Department of Physics, Columbia University, New York, NY 10027, USA
2Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA
3Department of Astronomy, Columbia University, New York, NY 10027, USA

(Dated: March 9, 2016)

Constraining cosmology using weak gravitational lensing consists of comparing a measured feature vector of dimension \(N_b\) with its simulated counterpart. An accurate estimate of the \(N_b \times N_b\) feature covariance matrix \(C\) is essential to obtain accurate parameter confidence intervals. When \(C\) is measured from a set of simulations, an important question is how large this set should be. To answer this question, we construct different ensembles of \(N_r\) realizations of the shear field, using a common randomization procedure that recycles the outputs from a smaller number \(N_s \leq N_r\) of independent ray-tracing \(N\)-body simulations. We study parameter confidence intervals as a function of \((N_s, N_r)\) in the range \(1 \leq N_s \leq 200\) and \(1 \leq N_r \lesssim 10^5\). Previous work [1] has shown that Gaussian noise in the feature vectors (from which the covariance is estimated) lead, at quadratic order, to an \(O(1/N_r)\) degradation of the parameter confidence intervals. Using a variety of lensing features measured in our simulations, including shear-shear power spectra and peak counts, we show that cubic and quartic covariance fluctuations lead to additional \(O(1/N_r^2)\) error degradation that is not negligible when \(N_r\) is only a factor of few larger than \(N_s\). We study the large \(N_r\) limit, and find that a single, 240Mpc/h sized \(512^3\)-particle \(N\)-body simulation \((N_s = 1)\) can be repeatedly recycled to produce as many as \(N_r = \text{few} \times 10^5\) shear maps whose power spectra and high-significance peak counts can be treated as statistically independent. As a result, a small number of simulations \((N_s = 1\ or\ 2)\) is sufficient to forecast parameter confidence intervals at percent accuracy.

PACS numbers: 98.80.-k, 95.36.+x, 95.30.Sf, 98.62.Sb
Keywords: Weak Gravitational Lensing — Simulations — Methods: analytical, numerical, statistical

I. INTRODUCTION

Weak gravitational lensing (WL) is a promising cosmological probe for constraining the dark energy equation of state \(w\), and has been considered by a range of past (CFHTLens [2, 3], COSMOS [4]), ongoing (DES [5]) and future (LSST [6], Euclid [7], WFIRST [8]) experiments. In an era where cosmology is data driven, accurate numerical simulations of shear fields are becoming important for several reasons, including assessing baryonic effects [9–14], the utility of non–Gaussian statistics [15–23] and various systematic effects [24–27].

A fundamental issue with predictions from simulations is that the finite number of simulations naturally introduces fluctuations in the forecasts, due to inevitable sample variance [28]. In general, quantities such as the mean or the variance of any feature (e.g. the shear power spectrum at a multipole \(l\)), measured from a finite set of simulations, will fluctuate, and can also suffer a bias. While biases in the estimates of both the mean and the variance have been studied extensively, the impact of fluctuations in the variance has received less attention. These fluctuations have been shown to have non-negligible effects on estimates of features covariances and hence on parameter constraints. In particular, in the limit of Gaussian fluctuations, the parameter confidence limits are degraded by a factor \(1 + O(1/N_r)\) [1, 22].

This work studies these issues further, focusing on the number of independent \(N\)-body simulations required for an accurate estimate of the parameter constraints. Ray-tracing simulations that resolve the cosmic structures responsible for lensing on arcminute scales are limited to physical sizes of hundreds of Mpc, and thus cover a solid angle of only \(O(10\ deg^2)\). As a result, many simulations are required to tile a significant fraction of the sky, and to make predictions for large “all-sky” surveys, such as the ones by DES, LSST, Euclid, WFIRST. In practice, this has led to the wide-spread use of “pseudo-independent” realizations, i.e. a procedure in which one randomizes and re-cycles the output of a single 3D simulation multiple times. In light of the forthcoming large surveys, it is imperative to assess the statistical validity of this approach, and to ask how many times a single simulation can be fairly recycled. In this paper, we address these questions with ensembles of up to \(N_r = 10^5\) random realizations, extracted from up to \(N_s = 200\) independent ray-tracing \(N\)-body simulations. We focus in particular on the parameter \(w\), and on two different statistics: the (convergence) power spectrum and the number counts of peaks.

This paper is organized as follows. In § II, we summarize the shear simulation methods we utilized, and describe the formalism we adopted to forecast cosmological parameter constraints. We then vary the number of simulations and the number of pseudo-independent realizations, and present our main findings in § III. These results are discussed further in § IV. We offer our conclusions, and suggest follow-up future work in § V.
II. METHODS

A. Ray-tracing simulations of the convergence field

In this section, we describe how we constructed our shear field ensembles. Background galaxies at redshift $z_s$ are lensed by large scale structures between $z = 0$ and $z_s$. The shape distortions due to the cosmic shear $\gamma$ can be computed in terms of the dark matter gravitational potential $\Phi(x, z)$. Because the evolution of $\Phi$ with redshift is non-linear, it needs to be computed with numerical simulations. We make use of the public code Gadget2 \cite{2005astro.ph..9573S}, with which we run a sequence of 200 independent $N$-body simulations that track the evolution of the density fluctuations. We assume a standard $\Lambda$CDM background universe with the parameters $(\Omega_m, \Omega_\Lambda, h, w, \sigma_8, n_s) = (0.26, 0.74, 0.72, -1, 0.8, 0.96)$. We fix the comoving size of the simulation box to 240 Mpc$/h$, and use 512$^3$ particles, corresponding to a dark matter particle mass of $\approx 10^{10} M_\odot$.

We assume a uniform galaxy distribution at a constant redshift $z_s = 2$ (at which the simulation box has an angular size of $\theta_{\text{box}} = 3.5^\circ$) and we discretize the mass distribution between $z_s$ and the observer at $z = 0$ with a sequence of 46 two dimensional lenses of thickness 80 Mpc$/h$. The surface density on each lens plane is computed by projecting the three-dimensional density distribution of a single $N$-body simulation box for each orientation $n_z, n_y, n_x$: the lens plane will be perpendicular to this direction.

Let \( \hat{d} \) be a single estimate for a feature of dimension $N_r$, $d(p)$ be the true value of this feature at a point $p$ in parameter space (which has a dimension $N_p$) and $C$ be the $N_b \times N_b$ feature covariance matrix. For the purpose of this work $p$ is the triplet $(\Omega_m, w, \sigma_8)$ and $d$ is one of the features – either a power spectrum or a peak count histogram – in Table I. Although existing emulators can be used, in principle, to compute both $d(p)$ and $C$, the latter is more difficult, and typically only the mean, $\mu(d(p))$, has been computed to date (refs. \cite{Carrick:2018wsr, Xue:2018qoz}, but see an exception by ref. \cite{2017MNRAS.470..699B}). Estimating $C$ from simulations involves generating a series of mock realizations $\hat{d}_r$ with $r = 1...N_r$ and computing the sample covariance $\hat{C}$,

\[
\hat{C} = \frac{1}{N_r} \sum_{r=1}^{N_r} \hat{d}_r \hat{d}_r^T.
\]
\[ \hat{C} = \frac{1}{N_r - 1} \sum_{r=1}^{N_r} (\tilde{d}_r - \tilde{d})(\tilde{d}_r - \tilde{d})^T. \]

Assuming a normal feature likelihood, together with a flat prior on the parameter space, the parameter posterior distribution \( \mathcal{L}(\mathbf{p} | \tilde{d}_{\text{obs}}) \) given an observed instance of \( \tilde{d} \), which we call \( \tilde{d}_{\text{obs}} \), follows from Bayes’ theorem,

\[ -2 \log \mathcal{L}(\mathbf{p} | \tilde{d}_{\text{obs}}) = | \tilde{d}_{\text{obs}} - \mathbf{d}(\mathbf{p}) |^T \hat{C}^{-1} [ \tilde{d}_{\text{obs}} - \mathbf{d}(\mathbf{p}) ]. \]

For the sake of simplicity, we approximate the posterior as a Gaussian around its maximum. This corresponds to Taylor-expanding the simulated feature to first order around a point \( \mathbf{p}_0 \) (ideally the maximum of Eq. 4):

\[ \mathbf{d}(\mathbf{p}) \approx \mathbf{d}_0 + \mathbf{d}_0'(\mathbf{p} - \mathbf{p}_0). \]

We chose \( \mathbf{p}_0 \) to be the triplet \((\Omega_m, w, \sigma_s) = (0.26, -1, 0.8)\). To measure the derivatives of the features \( \mathbf{d}_0' \) with respect to the cosmological parameters, we make use of the public code Nicasa \[33\] for the power spectrum, and we use an independent simulation set (containing simulations with a variety of different combinations of \((\Omega_m, w, \sigma_s)\), see Table 1 for the peak counts.

We can build the estimator for the posterior maximum \( \hat{\mathbf{p}} \), given the observation \( \tilde{d}_{\text{obs}} \), as follows:

\[ \hat{\mathbf{p}} = \mathbf{p}_0 + \hat{T}(\tilde{d}_{\text{obs}} - \mathbf{d}_0), \]

\[ \hat{T} = (\mathbf{d}_0' \hat{C}^{-1} \mathbf{d}_0')^{-1} \mathbf{d}_0' \hat{C}^{-1}. \]

Because \( \hat{\mathbf{p}} \) is estimated using a single noisy data instance \( \tilde{d}_{\text{obs}} \), its estimate will be scattered around the true value \( \langle \hat{\mathbf{p}} \rangle_\text{O} \). In the following we use the \( \langle \cdot \rangle_\text{O} \) notation for expectation values taken with respect to observations, while we keep the notation \( \langle \cdot \rangle \) for expectation values taken with respect to the simulations. Defining the precision matrix \( \hat{\Psi} = \hat{C}^{-1} \), we can express the estimator of the observational scatter in \( \hat{\mathbf{p}} \):

\[ \hat{\Sigma}_p = \hat{\mathbf{F}}^{-1} \mathbf{d}_0' \hat{\Psi} \mathbf{d}_0, \]

\[ \hat{\mathbf{F}} = \mathbf{d}_0'^T \hat{\Psi} \mathbf{d}_0. \]

Here we introduced the familiar Fisher matrix estimator \( \hat{\mathbf{F}} = \mathbf{d}_0'^T \hat{\Psi} \mathbf{d}_0 \) and, for simplicity, we assumed \( \langle \tilde{d}_{\text{obs}} \rangle_\text{O} = \mathbf{d}_0 \), so that \( \langle (\tilde{d}_{\text{obs}} - \mathbf{d}_0)(\tilde{d}_{\text{obs}} - \mathbf{d}_0)^T \rangle_\text{O} = \mathbf{C} \). When we perform an observation \( \tilde{d}_{\text{obs}} \), the parameter estimate \( \hat{\mathbf{p}} \) is a random draw from a probability distribution with variance \( \hat{\Sigma}_p \), which inherits noise from the simulations. The noise in the covariance estimator (Eq. 3) and in its inverse \( \hat{\Psi} \) propagate all the way to the posterior (Eq. 1), the parameter estimate (Eq. 6) and its variance (Eq. 8). Following refs. \[1, 29, 38\] we compute the expectation value of Eq. 8 over simulations, \( \langle \hat{\Sigma}_p \rangle \), up to \( O(1/N_r^2) \) by expanding Eq. 8 to quartic order in the statistical fluctuations of \( \hat{\Psi} \). Denoting the true parameter covariance, i.e. the usual inverse Fisher matrix, as \( \Sigma_p = \mathbf{F}^{-1} \), we find the result \[39, 40\]:

\[ \langle \hat{\Sigma}_p \rangle = \Sigma_p \left[ 1 + \frac{N_b - N_p}{N_r} + \frac{(N_b - N_p)(N_b - N_p + 2)}{N_r^2} \right] + O \left( \frac{1}{N_r^2} \right). \]

This is the result obtained by ref. \[1\].

2. Usually the true data covariance \( \mathbf{C} \) is unknown, and it is tempting to plug in its estimator \( \hat{\mathbf{C}} \), measured from the same simulation set we use to compute \( \hat{\Psi} \). This approach has been used before in the literature (e.g. \[17, 24\]). If this is done without correcting for the bias in \( \hat{\Psi} \) (see Ref. \[29\] and eq. \[20\] below), the parameter variance will have a contribution from both the second and first-order fluctuations in \( \hat{\Psi} \), which now have a nonzero expectation value. In
III. RESULTS

In this section we present the main results of this work. We show the qualitative behavior of a variety of feature \( \tilde{d}_r \) probability distribution functions (PDFs) in ensembles built with different \( N_s \) and \( N_r \). In Figure 1 we show the PDF of the power spectrum at four selected multipoles, spanning the linear (\( \ell = 115 \)) to the nonlinear (\( \ell = 5283 \)) regime. In Figure 2 we show the ensemble mean for these power spectra, as well as for peak counts of three different \( \kappa_0 \) heights (corresponding to \( \approx 2 \)–13\( \sigma \) peaks), as a function of \( N_s \). In Figure 3 we show the variance of the power spectrum at each multipole, as a function of \( N_s \), in units of the variance expected if the convergence \( \kappa \) was a Gaussian random field

\[
\text{Var}(P_{\ell}^{\kappa}) = \frac{(P_{\ell}^{\kappa})^2}{N_{\text{eff}}(\ell)}.
\]

Here \( N_{\text{eff}}(\ell) \) is the number of independent modes used to estimate the power spectrum at \( \ell \).

In practice, we measure \( P_{\ell}^{\kappa} \) on the Fourier transform of the pixelized simulated map \( \kappa_r(\theta) \), using the FFT algorithm, and some care must be taken to count the number of modes \( N_{\text{eff}}(\ell) \) correctly. Each pixel \( (i_x, j_y) \) in Fourier space corresponds to a mode \( (\ell_x, \ell_y) = 2\pi(i_x, i_y)/\theta_{\text{box}} \), with \( \ell_x = -n_{\text{ray}}/2, ..., n_{\text{ray}}/2 \) and \( i_y = 0, ..., n_{\text{ray}}/2 \). Here \( n_{\text{ray}} = 2048 \) is the linear number of pixels on the ray–traced convergence maps. We count the number of pixels \( N(\ell) \) that fall inside a multipole bin \( (\ell_1, \ell_2) \). Because the \( \kappa \) field is real, the modes \( (\pm \ell_x, 0) \) are not independent. If we let \( N(\ell, \ell_y = 0) \) be the number of non–independent modes, the effective number of independent modes for the variance is given by

\[
N_{\text{eff}}(\ell) = \frac{N^2(\ell)}{N(\ell) + N(\ell, \ell_y = 0)}.
\]

This correction is important at low \( \ell \), where pixelization effects are non-negligible; \( N_{\text{eff}}(\ell \gg 2\pi/\theta_{\text{box}}) \approx N(\ell) \).

In Figures 4 and 5 we show the dependence of the confidence range \( \langle \tilde{\sigma}_p^2 \rangle \) on \( N_r \), derived from the features used in this work (see Table III for a comprehensive list).

Table I. Summary of the shear ensembles used in this work. \( N_s \) and \( N_r \) refer to the number of independent \( N \)-body simulations, and the number of pseudo-independent realizations created from these simulations, respectively.

| \( \Omega_m \) | \( w \) | \( \sigma_\theta \) | \( \langle P_{\ell}^{\kappa} \rangle \) | \( (N_s, N_r) \) | Number of \( \kappa \) ensembles |
|---|---|---|---|---|---|
| 0.26 | -1 | 0.8 | 1 (to 200,1024) | 16 |
| 0.26 | -1 | 0.8 | 1 (128000) | 1 |
| 0.29 | -1 | 0.8 | 1 (1024) | 1 |
| 0.26 | -0.8 | 0.8 | 1 (1024) | 1 |
| 0.26 | -1 | 0.6 | 1 (1024) | 1 |

4
FIG. 1. PDF of the $\kappa$ power spectrum $\mathcal{L}(P_\kappa)$ at four selected multipoles $\ell = 115, 344, 1349, 5283$, for different shear ensembles constructed from on $N_s = 1$ (black), 2 (blue), 5 (green), 50 (red), and 100 (purple) independent $N$–body simulations. Each curve is based on $N_r = 1024$ realizations. The dashed black curves correspond to ensembles generated with $N_s = 1$ and $N_r = 128000$. For $N_s \geq 2$, the distributions appear similar to the eye; this similarity is confirmed by the comparisons in Figures 2 and 3 below.

| Feature                  | Specifications | $N_b$ | Symbol | Color |
|--------------------------|----------------|-------|--------|-------|
| Power Spectrum, log binning | $\ell \in [100, 800]$ | 8     | $\times$ | black |
| Power Spectrum, log binning | $\ell \in [1000, 6000]$ | 7     | ■ | black |
| Power Spectrum, log binning | $\ell \in [100, 6000]$ | 15    | ● | red |
| Power Spectrum, linear binning | $\ell \in [100, 2000]$ | 15    | $+$ | red |
| Power Spectrum, linear binning | $\ell \in [2500, 4500]$ | 15    | $\times$ | red |
| Power Spectrum, linear binning | $\ell \in [100, 4500]$ | 30    | ● | green |
| Power Spectrum, linear binning | $\ell \in [100, 6000]$ | 39    | ● | blue |
| Low peaks               | $\kappa_0 \in [-0.06, 0.09]$ | 15    | $+$ | red |
| Intermediate peaks       | $\kappa_0 \in [0.1, 0.27]$ | 15    | $\star$ | red |
| High peaks               | $\kappa_0 \in [0.28, 0.45]$ | 15    | $\diamond$ | red |
| Low+Intermediate peaks   | $\kappa_0 \in [-0.06, 0.27]$ | 30    | $\times$ | green |
| Intermediate+High peaks  | $\kappa_0 \in [0.1, 0.45]$ | 30    | ■ | green |
| All peaks                | $\kappa_0 \in [-0.06, 0.45]$ | 45    | ■ | magenta |

TABLE II. Catalog of feature types used in this work, along with the chosen number of bands $N_b$ and the plot legends for Figure 4.
measured directly from the simulations with the analytic
from equation (16). The variance is measured from different
\( \ell \)
the multipole
\( \kappa 
width for the peak counts has been fixed to \( \Delta \kappa = 0.011 \). The dashed black line shows a level of 0.1\( \sigma \) accuracy for reference. For \( N_s \geq 2 \), the means are statistically indistinguishable (even at \( \sim 0.1\sigma \)) from those in the ensemble with \( N_s = 200 \).

large \( N_s \), trends of the \( w \) constraint. Figure 4 illustrates the behavior at relatively low \( N_r \), and compares \( \langle \sigma^2_w \rangle \) measured directly from the simulations with the analytic expectations from equation (16). Finally, in Figure 5 we show how the \( w \) confidence limit changes with \( N_s \).

FIG. 2. The mean value \( \bar{d} \) of various features, measured from ensembles created from different numbers \( N_s \) of simulations. For each case, the difference compared to the mean in the \( N_s = 200 \) ensemble is shown, in units of the statistical error measured in the \( N_s = 200 \) ensemble. The colored curves refer to shear–shear power spectra measured at \( \ell = 115 \) (black), 1027 (cyan), and 5283 (green), and peak counts with heights \( \kappa_0 = 0.05 \) (red), 0.17 (purple), and 0.28 (orange). The \( \kappa \) bin width for the peak counts has been fixed to \( \Delta \kappa = 0.011 \). The dashed black line shows a level of 0.1\( \sigma \) accuracy for reference. For \( N_s \geq 2 \), the means are statistically indistinguishable (even at \( \sim 0.1\sigma \)) from those in the ensemble with \( N_s = 200 \).

FIG. 3. Variance of the \( \kappa \) power spectrum as a function of the multipole \( \ell \), in units of the expected Gaussian variance from equation (16). The variance is measured from different shear ensembles based on \( N_s = 1 \) (black), 2 (blue), 5 (green), 10 (red), 50 (purple), or 100 (orange) N-body simulations. Non-Gaussianities of the underlying structures increase the variance on small scales, but no clear trend with \( N_s \) can be identified on any scale.

FIG. 4. Expectation value of the variance of \( w \) computed from equation (8), shown as a function of \( 1/N_r \). The different symbols and colors correspond to the features listed in Table II. The dashed and solid curves show the analytic predictions from equation (10) at orders \( O(1/N_r) \) and \( O(1/N_r^2) \), respectively. The thick solid curves show the empirical predictions from equation (11). The asymptotic variance \( \sigma_{w, \infty}^2 \) has been computed from a linear regression of \( \langle \sigma_w^2 \rangle \) vs \( 1/N_r \) for \( N_r > 500 \). The figure clearly shows that terms beyond \( O(1/N_r) \) need to be considered, unless \( N_r \gg N_s \).

FIG. 5. Bias in the variance of \( w \), \( \langle \sigma^2_w \rangle - \sigma_{w, \infty}^2 \), as a function of the number of realizations \( N_r \) used to estimate the covariance (eq. 3). The figure shows both the trend measured in the simulations (solid lines) and their scaling expected from Eq. (16) with \( D = N_s - N_b \) (dashed line). The asymptotic variance \( \sigma_{w, \infty}^2 \) has been estimated to be the value \( \langle \sigma_w^2 \rangle (N_r = 10^3) \). Different features are considered: power spectra with logarithmically spaced \( \ell \) in \([100, 6000]\) (black), \( \ell \in [100, 250]\) (red), peak counts in the unsmoothed maps with height \( \kappa_0 \in [0.44, 0.48] \) (green) and peak counts in the smoothed maps (with a Gaussian kernel of size \( \theta_G = 1' \) ) with height \( \kappa_0 > 0.15 \) (blue). No deviations from the expected \( 1/N_r \) behavior are observed up to \( N_r \approx 0.01 \times 10^5 \), except for the large-scale power spectrum, in which case the deviations occur much earlier \( (N_r \approx 10^3) \).
investigates the parameter errors. This figure shows that when we infer the variance of convergence power spectrum ensemble means. We at-tribute these offsets to large (random) statistical errors. As pointed out by [42], the box size used for the N-body runs can be recycled repeatedly. However, it is not sufficient by itself to conclude that N does not impact the parameter inferences, since these depend on the cross band covariances.

Figure 4 investigates the parameter errors. This figure shows that error degradation estimates truncated at order O(1/Nr) are too optimistic when the number of simulations Nr used to measure the covariance is only a factor of few larger than the dimension of the feature space Ns.

In these cases effects coming the next–to–leading orders O(1/Nr²) become non–negligible on constraint degradation. In particular, we find that already for Ns = 30 and Nr ∼ 100, the error degradation estimates to the next leading order, O(1/Nr³), remain too optimistic. Accurate analytic estimates in this regime would require at least terms of order O(1/Nr³), which come from higher–than–quartic fluctuations.

In Figure 6, we examine how the degradation in the w constraint depends on the number of simulations used to estimate the covariance, in the limit of large Nr. We find excellent agreement with the expected scaling (eq. 15) up to Nr ∼ few × 10⁴ when using the κ power spectrum in the multipole range ℓ ∈ [100, 6000]. The same behavior is observed when considering the high-significance peak counts (> 10σ for unsmoothed maps and > 5σ for 1’ smoothed maps). As the figure shows, around these values of Nr, the (σw²) − σw,∞ curve becomes noisy and reaches negative values. This is a clear indication that the 1/Nr behavior is broken and a plateau in (σw²) is reached. The negative values in the plot are a consequence of the noise in the estimation of this plateau value (or equivalently in the estimated value of σw,∞).

We conclude that a single N-body simulation is sufficient to construct an ensemble of up to a few × 10⁴ mutually independent convergence power spectra. For Nr ≫ 10⁴, the shear realizations can no longer be considered independent. We emphasize that the precise value of this Nr will depend on the size of the simulation box (which, in our case, is (240 Mpc/h)³, with 512³ particles) and also on the range of multipoles ℓ used to constrain the parameters. Figure 5 shows that when we infer w only from large–scale modes, ℓ ≤ 250, the plateau is reached at least an order of magnitude earlier in the number of realizations. In other words, the number of independent power spectra we can generate decreases as we increase the spatial scales of interest. This is due to the fact that, because of the finite box size, the number of independent lens plane shifts (as described in § 11A) decreases as the mode size approaches the size of the box. Similarly, one may expect that the independence in the statistics of high-amplitude peaks, which are predomi-
nantaly produced by single massive halos, may be compromised by these halos being present repeatedly, in many of the pseudo-independent realizations. However, Figure 6 shows that this is not the case: the peak count statistics are shown at $\kappa$ thresholds corresponding to massive $(\approx 10^{15} \, M_\odot)$ halos, yet there is no evidence that the independence of the maps breaks down until $N_r \rightarrow \infty$. Apparently, randomly projected structures, which vary from realization-to-realization, contribute significantly to the statistics of these high peaks.

Figure 6 shows how the “true” $w$ constraint (in the limit $N_r \rightarrow \infty$; or equivalently the $w$ constraint with the known $N_r$ dependence factored out), depends on $N_r$. We find that, in the range $N_r \in [1, 200]$ the inferred $w$–variance $\sigma^2_{w, \infty}$ fluctuates stochastically only by 1%, and does not show any trend with $N_r$.

Finally, we found that when we estimate the data covariance $\mathbf{C}$ from the same simulation set used to measure $\hat{\Psi}$, the effective dimensionality $D$ decreases with increasing $N_r$ in the case where the $\hat{\Psi}$ bias is not corrected (eq. 13). This $N_r$-dependence disappears when the bias is corrected (eq. 14). This fact that should be taken into consideration when forecasting parameter errors purely from simulations, as the errors will otherwise be underestimated. A similar conclusion was reached by [14] (although their paper did not address the impact of using the same simulation set for $\mathbf{C}$ and $\hat{\Psi}$).

V. CONCLUSIONS

In this work, we have examined the effect of forecasting cosmological constraints based on shear ensembles generated from a finite number of $N$–body simulations. Our main results can be summarized as follows:

- When the feature covariance matrix is measured from simulations, parameter constraints are degraded. This degradation is appreciably larger than the $O(1/N_s)$ computed by [1] when the number of realizations $N_s$ is only a factor of few larger than the feature vector size $N_h$.
- We can recycle a single $240\text{Mpc}/h$ $N$–body simulation to produce an ensemble of $O(10^4)$ shear maps whose small-scale power spectra and high-significance peak counts are statistically independent. The mean feature measured from a shear ensemble, though, could be inaccurate if only one $N$–body simulation is used.
- As few as one or two independent $N$–body simulations are sufficient to forecast $w$ error bars to 1% accuracy, provided that a sufficiently large number $N_r$ of realizations are used to measure feature covariances. In particular, provided that biases in the inverse covariance are corrected, percent–level forecasts require $N_r \gtrsim 100(N_h - N_r)$ realizations.
- Depending on the feature type used to constrain cosmology, a larger number of $N$–body simulations might be needed to measure accurate ensemble means to an accuracy corresponding 10% of the statistical error. If this accuracy requirement is relaxed to 50% of the statistical error, we find that as low as $N_s = 2$ simulations are sufficient for the feature types we consider in this work.

Future extensions of this work should involve extending our analysis to a larger set of cosmological parameters, and to more general feature spaces, such as the ones that characterize non–Gaussian statistics (e.g. including higher moments of the $\kappa$ field, Minkowski Functionals, and higher-order $\kappa$ correlators). While our results are highly encouraging, and suggest that a single $N$–body simulation can be recycled repeatedly, to produce as many as $10^4$ independent shear power spectra or peak count histograms. In order to scale our results to large future surveys, such as LSST, it will be necessary to determine if our findings hold when challenged by larger and higher-resolution $N$–body simulations [45].

ACKNOWLEDGEMENTS

We thank Lam Hui for useful discussions. The simulations in this work were performed at the NSF XSEDE facility, supported by grant number ACI-1053575, and at the New York Center for Computational Sciences, a cooperative effort between Brookhaven National Laboratory and Stony Brook University, supported in part by the State of New York. This work was supported in part by the U.S. Department of Energy under Contract Nos. DE-AC02-98CH10886 and DE-SC00012704, and by the NSF Grant No. AST-1210877 (to Z.H.) and by the Research Opportunities and Approaches to Data Science (ROADS) program at the Institute for Data Sciences and Engineering at Columbia University (to Z.H.).

[1] S. Dodelson and M. D. Schneider, Phys. Rev. D 88, 063537 (2013) arXiv:1304.2593 [astro-ph.CO]
[2] C. Heymans, L. Van Waerbeke, L. Miller, T. Erben, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mel-
Sample variance is a broad term that has been used in the literature to describe a range of phenomena. Throughout this paper, we use it to refer to the fluctuations in an ensemble of simulations, which represent random realizations of the same initial conditions.

L. Miller, C. Heymans, T. D. Kitching, L. van Waerbeke, T. Erben, H. Hildebrandt, H. Hoekstra, Y. Mellier, B. T. P. Rowe, J. Coupon, J. P. Dietrich, L. Fu, J. Harnois-Déрапs, M. J. Hudson, M. Kilbinger, K. Kuijken, T. Schrabback, E. Semboloni, S. Vafaei, and M. Vlendar, MNRAS 429, 2858 (2013)
arXiv:1210.8201 [astro-ph.CO]

J. Amiaux, R. Scaramella, Y. Mellier, B. Altieri, C. Burigana, A. Da Silva, P. Gomez, J. Hoar, R. Laureijs, E. Maioaran, D. Magalhães Oliveira, F. Renk, G. Saavedra Criado, I. Tereno, J. L. Auguères, J. Brinchmann, M. Cropper, L. Duvet, A. Ealet, P. Franzetti, B. Garilli, P. Gondoin, L. Guzzo, H. Hoekstra, R. Holmes, K. Jahnke, T. Kitching, M. Meneghetti, W. Percival, and S. Warren, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8442 (2012) p. 84420Z, arXiv:1209.2228 [astro-ph.IM]

J. Xiang, J. M. Kratochvil, K. Huffenberger, Z. Haiman, and M. May, Phys. Rev. D 87, 023511 (2013) arXiv:1210.0608

E. Semboloni, H. Hoekstra, and J. Schaye, MNRAS 434, 148 (2013) arXiv:1210.7303

M. White, Astrophysical Space Science 22, 211 (2004) astro-ph/0405593

H. Zhan and L. Knox, ApJ 616, L75 (2004) astro-ph/0409198

A. R. Zentner, D. H. Rudd, and W. H. Hu, Phys. Rev. D 77, 043507 (2008)

D. H. Rudd, A. R. Zentner, and A. V. Kravtsov, ApJ 672, 19 (2008) astro-ph/0703741

X. Yang, J. M. Kratochvil, S. Wang, E. A. Lim, Z. Haiman, and M. May, Phys. Rev. D 84, 043529 (2011) arXiv:1109.6333

J. M. Kratochvil, E. A. Lim, S. Wang, Z. Haiman, M. May, and K. Huffenberger, Phys. Rev. D 85, 103513 (2012) arXiv:1109.6334 [astro-ph.CO]

A. Petri, Z. Haiman, L. Hui, M. May, and J. M. Kratochvil, Phys. Rev. D 88, 123002 (2013) arXiv:1309.4460

L. Marian, R. E. Smith, and G. M. Bernstein, ApJ 698, L33 (2009) arXiv:0811.1991

M. Takada and B. Jain, MNRAS 337, 875 (2002) astro-ph/0205055

M. Takada and B. Jain, MNRAS 344, 857 (2003) astro-ph/0304034

M. Takada and B. Jain, MNRAS 348, 897 (2004) astro-ph/0310125

J. Bergé, A. Amara, and A. Réfrégier, ApJ 712, 992 (2010) arXiv:0909.0529

J. P. Dietrich and J. Hartlap, MNRAS 402, 1049 (2010) arXiv:0906.3512

M. Shirasaki, N. Yoshida, and T. Hamana, ApJ 774, 111 (2013) arXiv:1304.2164 [astro-ph.CO]

D. Bard, J. M. Kratochvil, C. Chang, M. May, S. M. Kahn, Y. AlSayyad, Z. Ahmad, J. Bankert, A. Connolly, R. R. Gibson, K. Gilmore, E. Grace, Z. Haiman, M. Hannel, K. M. Huffenberger, J. G. Jernigan, L. Jones, S. Krughoff, S. Lorenz, S. Marshall, A. Meert, S. Nagarajan, E. Peng, J. Peterson, A. P. Rasmussen, M. Shmukler, N. Sylvestre, N. Todd, and M. Young, MNRAS 428, 2695 (2013) arXiv:1206.1378

D. Huterer, M. Takada, G. Bernstein, and B. Jain, MNRAS 366, 101 (2006) astro-ph/0506030

Sample variance is a broad term that has been used in the literature to describe a range of phenomena. Throughout this paper, we use it to refer to the fluctuations in an ensemble of simulations, which represent random realizations of the same initial conditions.

A. Taylor, B. Joachimi, and T. Kitching, MNRAS 432, 1928 (2013) arXiv:1212.4359

V. Springel, MNRAS 364, 1105 (2005) astro-ph/0505010

S. Hilbert, J. Hartlap, S. D. M. White, and P. Schneider, A&A 499, 31 (2009) arXiv:0809.5035

B. Jain, U. Seljak, and S. White, ApJ 530, 547 (2000) astro-ph/9901191

A. Petri, “The lensools computing package.” (2015).

M. Sato, T. Hamana, R. Takahashi, M. Takada, N. Yoshida, T. Matsubara, and N. Sugiyama, ApJ 701, 945 (2009) arXiv:0906.2237 [astro-ph.CO]

K. Heitmann, D. Higdon, M. White, S. Habib, B. J. Williams, E. Lawrence, and C. Wagner, ApJ 705, 156 (2009) arXiv:0902.0429 [astro-ph.CO]

M. Kilbinger, K. Benabed, J. Guy, P. Astier, I. Tereno, L. Fu, D. Wraith, J. Coupon, Y. Mellier, C. Ballard, F. R. Bouchet, T. Hamana, D. Hardin, H. J. McCracken, R. Pain, N. Regnault, M. Schuhtheis, and H. Yahagi, A&A 497, 677 (2009) arXiv:0810.5129

M. D. Schneider, L. Knox, S. Habib, K. Heitmann, D. Higdon, and C. Nakhlé, Phys. Rev. D 78, 063529 (2008)

S. Matsumoto, ArXiv e-prints (2010), arXiv:1004.4717 [math.ST]

The details of the calculation are given in Appendix A.
APPENDIX A: CUBIC AND QUARTIC COVARIANCE FLUCTUATIONS

The goal of this appendix is to give a derivation of eq. (11). When the simulated feature vector $\mathbf{d}_r$ is drawn from a Gaussian distribution, the covariance estimator $\mathbf{C}$ follows the Wishart distribution, and its inverse $\hat{\mathbf{C}}$ follows the inverse Wishart distribution (see Ref. [29] for analytical expressions for these probability distributions). Computing expectation values of eq. (8) over the inverse Wishart distribution is not possible analytically, and a perturbative expansion is necessary. Writing $\hat{\mathbf{C}} = \mathbf{C} + \delta \hat{\mathbf{C}}$, we can expand eq. (8) in powers of $\delta \hat{\mathbf{C}}$.

The expectation value of each term in this expansion can be calculated in terms of moments of the inverse Wishart distribution. Ref. [38] provides a general framework to compute these moments, and give exact expressions for moments up to quartic order. First, let us expand the inverse of the Fisher matrix estimator (eq. (9) in powers of $\delta \hat{\mathbf{C}}$). The $n$–th order of this expansion will be

$$\delta \hat{\mathbf{F}}^{-1}_n = (-1)^n (\mathbf{F}^{-1} \delta \hat{\mathbf{F}})^n \mathbf{F}^{-1}$$

with

$$\delta \hat{\mathbf{F}} = \mathbf{d}_0^T \delta \hat{\mathbf{C}} \mathbf{d}_0$$

Using eq. (10), we can expand eq. (8) to an arbitrary order in $\delta \hat{\mathbf{C}}$, take the expectation values of the fluctuations over the inverse Wishart distribution, and finally arrive at eq. (11). We use the notation $\nu \equiv N_r - 1$ and $\gamma \equiv (\nu - N_b - 1)/2$, and we indicate with capital letters pairs of matrix indices, for example $I = (i_1, i_2)$, where $i_a = 1..N_b$. The main results we utilize from ref. [38] regarding the first four moments are (up to order $O(1/\nu^2)$)

$$\langle \hat{\Psi}_I \rangle = \frac{\nu}{2\gamma} \Psi_I$$

$$\langle \delta \hat{\Psi}_I \delta \hat{\Psi}_J \rangle = \frac{\nu^2 \gamma \Psi_I \Psi_J + \nu^2 \gamma \Psi_I \Psi_J}{4\gamma^2(\gamma - 1)(2\gamma + 1)}$$

$$\langle \delta \hat{\Psi}_I \delta \hat{\Psi}_J \delta \hat{\Psi}_K \rangle = \frac{\nu^3 \Psi_I \Psi_J \Psi_K}{8\gamma(\gamma - 1)(\gamma - 2)(\gamma + 1)(2\gamma + 1)}$$

$$\langle \delta \hat{\Psi}_I \delta \hat{\Psi}_J \delta \hat{\Psi}_K \delta \hat{\Psi}_L \rangle = \frac{\nu^4 (2\gamma^2 - 5\gamma + 9) \Psi_I \Psi_J \Psi_K \Psi_L}{16\gamma(\gamma - 1)(\gamma - 2)(\gamma - 3)(\gamma - 1)(\gamma + 1)(2\gamma + 1)(2\gamma + 3)}.$$  

Here the curly bracket notation is a shorthand for a symmetrization over pair of indices: for example

$$\Psi_{(I \Psi_J)} = \Psi_{i_1j_1} \Psi_{i_2j_2} + \Psi_{i_1j_2} \Psi_{i_2j_1}.$$  

Eq. (20) expresses the bias in the $\hat{\mathbf{C}}$ estimator that already appears in the literature [44]. If we want to use the bias-corrected $\hat{\mathbf{C}}$ estimator (required for the perturbative expansion of eq. [8]), we need to apply an additional factor of $(2\gamma/\nu)^n$ to eqs. (20–23), where $n$ is the order of the moment up to which we are applying the correction. If we limit ourselves to computing the expectation value of eq. (8) up to order $O(1/\nu^2)$, we do not need to worry about this correction for eqs. (22–23), as the dominant term here is already $O(1/\nu^2)$. The next step is expanding eq. (8) in powers of $\delta \hat{\mathbf{C}}$ up to fourth order: this is easily done:

$$\hat{\Sigma}_\mathbf{p} = \left( \mathbf{F}^{-1} + \sum_{n=1}^4 \delta \hat{\mathbf{F}}^{-1}_n \right) \mathbf{d}_0^T \langle \Psi + \delta \hat{\mathbf{C}} \rangle \mathbf{C} \langle \Psi + \delta \hat{\mathbf{C}} \rangle \mathbf{d}_0 \left( \mathbf{F}^{-1} + \sum_{n=1}^4 \delta \hat{\mathbf{F}}^{-1}_n \right).$$  

Carrying out the calculations is simpler than it looks: because of the structure of eq. (25), each term in the expansion is proportional to $\Sigma_{\mathbf{p}} f_{\alpha}(N_b, N_p) / N_c$, where
\( f_{a}(N_b, N_p) \) is a polynomial in \( N_b \) and \( N_p \). Terms proportional to \( N_b \) arise from index contractions of type \( \text{tr}(\Psi C) = N_b \), which come from symmetrization terms of type \( \Psi_{(i} \Psi_{j)} \). Symmetrization factors of type \( \Psi_{(i} \Psi_{j} \Psi_{k)} \) or \( \Psi_{(i} \Psi_{j} \Psi_{k} \Psi_{l)} \), on the other hand, give rise to contractions of type \( \text{tr}(\Psi C) \text{tr}(FF^{-1}) = N_b N_p \). Moreover, we know that, at every order \( O(1/N_r^a) \), \( f_{a}(N_b, N_p) \) has to be proportional to \( N_b - N_p \), because it must vanish when \( N_b = N_p \). The reason for this is that if the feature derivative matrix \( d'_0 \) is square and invertible (which it should be in absence of degeneracies), then eq. (8) reduces to

\[
\hat{\Sigma}_p = (d'_0)^{-1} C(d'^0_0)^{-1}.
\] (26)

Every trace of the noise is gone, hence powers of \( 1/N_r^a \) must not appear at any order if \( N_b = N_p \). Armed with the knowledge of the above considerations, we can compute the expectation value of eq. (25) at second, third and fourth order in \( \delta \hat{\Psi} \), keeping the terms that are at most \( O(1/\nu^2) = O(1/N_r^2) \). When the combinatorial factors that arise from the expansion of eq. (25) are properly computed and the expectation values over the inverse Wishart distribution are taken according to eqs. (20-23), the results take the form

\[
\begin{align*}
(\delta \hat{\Psi})^2 &\rightarrow \Sigma_p \frac{\gamma(N_b - N_p)}{(\gamma - 1)(2\gamma + 1)} = \Sigma_p \left[ \frac{N_b - N_p}{N_r} + \frac{(N_b - N_p)(N_b + 3)}{N_r^2} \right] \\
(\delta \hat{\Psi})^3 &\rightarrow -4\Sigma_p \frac{(N_b - N_p)(1 + N_p)}{N_r^2} \\
(\delta \hat{\Psi})^4 &\rightarrow 3\Sigma_p \frac{(N_b - N_p)(1 + N_p)}{N_r^2}.
\end{align*}
\] (27)

When the results from eq. (27) are summed, eq. (10) immediately follows.
APPENDIX B: NEGATIVE EFFECTIVE DIMENSIONALITY

The goal of this appendix is to give a justification for why the effective dimensionality \( D \) that appears in eq. (15) can be negative in some cases. When we use the same simulation set to estimate \( C, \hat{\Psi} \), eq. (8) reduces to the inverse Fisher estimator \( \hat{\Sigma} = \hat{F}^{-1} \). At second order in the \( \Psi \) fluctuations this becomes

\[
\hat{\Sigma}_p = F^{-1} + F^{-1} \left( -\delta \hat{F} + \delta \hat{F} F^{-1} \delta \hat{F} \right) F^{-1}
\]

If the biased estimator for \( \hat{\Psi} \) is used, we can use eqs. (20–21) at order \( O(1/\nu) \) to compute

\[
\langle \hat{\Sigma}_p \rangle = \Sigma_p \left( 1 - \frac{N_b + 1}{N_r} + \frac{1 + N_p}{N_r} \right) = \Sigma_p \left( 1 + \frac{N_p - N_b}{N_r} \right).
\]

We immediately see that the coefficient of \( 1/N_r \) is negative, because \( N_b > N_p \). This is the result shown in eq. (13). If the bias correction for \( \hat{\Psi} \) is applied, the first order terms \( \delta \hat{F} \) average to 0, and we are left with only the last term in the sum eq. (29), which immediately yields eq. (14).