Complementary cosmological simulations

Gábor Rácz1,2, Alina Kiessling1, István Csabai2, and István Szapudi3

1 Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
e-mail: gabor.racz@jpl.nasa.gov
2 Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Pf. 32, 1518 Budapest, Hungary
3 Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

Received 13 October 2022 / Accepted 30 January 2023

ABSTRACT

Context. Cosmic variance limits the accuracy of cosmological N-body simulations, introducing bias in statistics such as the power spectrum, halo mass function, or the cosmic shear.

Aims. We provide new methods to measure and reduce the effect of cosmic variance in existing and new simulations.

Methods. We ran pairs of simulations using phase-shifted initial conditions with matching amplitudes. We set the initial amplitudes of the Fourier modes to ensure that the average power spectrum of the pair is equal to the cosmic mean power spectrum from linear theory.

Results. The average power spectrum of a pair of such simulations remains consistent with the estimated nonlinear spectra of the state-of-the-art methods even at late times. We also show that the effect of cosmic variance on any analysis involving a cosmological simulation can be estimated using the complementary pair of the original simulation. To demonstrate the effectiveness of our novel technique, we simulated a complementary pair of the original Millennium run and quantified the degree to which cosmic variance affected its the power spectrum. The average power spectrum of the original and complementary Millennium simulation was able to directly resolve the baryon acoustic oscillation features.

Key words. large-scale structure of Universe – dark matter – methods: numerical

1. Introduction

Large-scale galaxy surveys are important probes to test the standard Lambda cold dark matter (ΛCDM) model of cosmology in addition to alternative cosmologies. Past surveys, such as the Automated Plate Measurement (APM) galaxy survey (Maddox et al. 1990; Baugh & Efstathiou 1994), 2dF (Percival et al. 2001), Two Micron All-Sky Survey (2MASS; Allgood et al. 2001), Sloan Digital Sky Survey (SDSS; Tegmark et al. 2004), Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005), Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2016), and the Panoramic Survey Telescope and Rapid Response System (PanSTARRS; Chambers et al. 2016), have provided valuable data regarding the evolution of the matter in our Universe. The upcoming Rubin (LSST Science Collaboration 2009), Roman (Green et al. 2012), Spectro-Photometer for the History of the Universe, Epoch of Reionization, and ices Explorer (SPHEREx) (Doré et al. 2014) and Euclid (Tutusaus et al. 2020) surveys will greatly extend these observations by mapping unprecedented volumes with previously unseen etendue. Constraints on cosmological parameters are derived by comparing observations and theoretical predictions using statistical quantities such as the power spectrum, halo mass function, or the cosmic-shear two-point correlation function. Cosmological N-body simulations are widely used for calculating these quantities at late times (e.g., Valdarnini & Borgani 1991; Rimes & Hamilton 2006; Jenkins et al. 2001; Casarini et al. 2012). To answer the demands of the large surveys, simulations were run to calculate the nonlinear evolution of the 3D matter distribution inside large volumes with high precision over the years, such as the Millennium run (Springel et al. 2005), Millennium-II (Boylan-Kolchin et al. 2009), Bolshoi (Klypin et al. 2011), Millennium-XXL (Angulo et al. 2012), MultiDark (Klypin et al. 2016), the Euclid flagship simulation (Potter et al. 2017), and the Outer Rim Simulation (Heitmann et al. 2019). The results of these simulations have been used in hundreds of research projects and are still regularly used today.

Both observations and cosmological simulations are affected by cosmic variance, because they are both only sampling a finite volume of the cosmic density field (Driver & Robotham 2010; Moster et al. 2011; Schneider et al. 2016). The most straightforward method to estimate the effects of cosmic variance is to use an ensemble of cosmological simulations such as the LasDamas (McBride et al. 2009), the Indra (Falck et al. 2021) or the ABACUSSUMMIT (Maksimova et al. 2021) simulation suites. By comparing the results of hundreds of simulations, the cosmic mean and variance can be directly calculated for a given quantity. Paired (Pontzen et al. 2016) and paired-and-fixed (Angulo & Pontzen 2016) initial conditions can effectively reduce the cosmic variance in simulations. Detailed descriptions of these initial conditions are presented in Sects. 2.1 and 2.2, respectively. By running just two simulations, the results of the paired-and-fixed initial conditions method can closely match the average of the ensembles of 300 simulations for density distribution, power spectrum, and bispectrum. Although the statistics of the average dark matter clustering are precise in these simulations, the covariances from paired-and-fixed simulations are suppressed, and as a consequence, these cannot be used for generating mock galaxy catalogs as shown in Klypin et al. (2020). Harnois-Déraps et al. (2019) pointed out that reducing
the effects of cosmic variance on the cosmic shear covariance calculation is possible with matched pairs of simulations.

Our aim is to extend the techniques shown above by proposing alternative ways to generate initial conditions for cosmological simulations. The outline of this paper is as follows: In Sect. 2, we overview the traditional algorithm used to generate initial conditions and propose modifications to this in order to reduce the cosmic variance. Then, in Sect. 3, we compare the different simulations that started from the modified initial conditions. In Sect. 4, we demonstrate our new best performing method on the original Millennium simulation by running its complementary pair. Using this new simulation, we estimate the effect of cosmic variance on the original Millennium power spectrum. Finally, we summarize our results.

2. Initial conditions for cosmological simulations

The main goal of initial-condition creation is to represent a \( \delta(x) = \rho(x)/\bar{\rho} - 1 \) overdensity field with discrete particles. For cosmological simulations, this field should be consistent with the \( P_{\text{target}}(k) \) initial power spectrum. The first step in this process is to generate the Fourier transform of the overdensity field (Hockney & Eastwood 1988; Sirko 2005):

\[
\delta(k) = \int \delta(x) e^{-i k x} d^3 x.
\]

As this is a complex field, it can be written as

\[
\delta(k) = \langle A(k) \rangle + i \cdot \langle B(k) \rangle = A(k) \cdot e^{i \phi(k)},
\]

and it should satisfy the usual Hermitian constraints, because its inverse Fourier transform is a real field. For a random Gaussian realization of a target \( P_{\text{target}}(k) \) initial power spectrum, \( a(k) \) and \( b(k) \) should be drawn independently from a Gaussian distribution (Klypin & Holtzman 1997)

\[
P_Q(a) da = \frac{1}{\sigma(k) \sqrt{2 \pi}} e^{-\frac{a^2}{2 \sigma(k)^2}} da,
\]

for every resolved \( k \) wavenumber vector. Equivalent to this is to generate the magnitude \( A(k) \) from the Rayleigh distribution

\[
P_R(A) dA = \frac{A}{\sigma(k)^2} e^{-\frac{A^2}{2 \sigma(k)^2}} dA,
\]

and the corresponding phase \( \phi(k) \) is chosen randomly from a uniform distribution on a range between 0 and \( 2 \pi \) independently (Sirko 2005), where \( \sigma(k)^2 = VP_{\text{target}}(k) / (2 \pi)^3 \), \( V \) is the volume of the simulation, and \( k = |k| \). After the initial field is generated in Fourier space, the \( \delta(x) \) overdensity field is calculated using the inverse Fourier transform. In the last step of the initial-condition generation, this Eulerian density is transformed into a Lagrangian representation of particles using the Zel’dovich approximation (Zel’dovich 1970) or the second-order Lagrangian perturbation theory (2LPT; Crocce et al. 2006; Jenkins 2010). The power spectrum of the traditional initial conditions,

\[
P_{\text{IC}}(k_1) = \frac{(2\pi)^3}{V} \delta_{k_1} (\delta(k_1) \delta(k_2)) = \frac{(2\pi)^3}{V} \langle A(k)^2 \rangle_{k_1 k_2},
\]

will not be equal to \( P_{\text{target}}(k) \), because the initial amplitudes are drawn from a Rayleigh distribution, and this is the main source of the cosmic variance in simulations. In this article, we use \( \langle A \rangle_{k_1 k_2} \) to denote the average value of the \( |k'| = k \) surface in Fourier space. In the rest of this paper, we discuss modifications to this traditional method and show how these modifications can reduce the effects of cosmic variance in cosmological \( N \)-body simulations.

2.1. Inverted initial conditions

This method was proposed by Pontzen et al. (2016) and can be used to create a density-inverted counterpart to an existing initial condition. This can be achieved by running the same initial condition generator again with the same parameters and random seed, but with \( \phi(k) \) phases shifted by \( \pi \). As a consequence of this phase shift, the initial \( \delta(x) \) overdensity fields are inverted in the counterpart, that is, underdensities are substituted for overdensities and vice versa. The simulations that start from the pairs of these initial conditions are called Paired simulations. The power spectrum of the original and the inverted initial condition are the same, because they only differ in the phases. Despite the fact that the pair have the same initial power spectrum, the cosmic variance can still be reduced by averaging the results of paired simulations, as the averaging cancels phase correlations that emerge from the late-time nonlinear evolution of the density field (Villaescusa-Navarro et al. 2018).

2.2. Fixed and paired-and-fixed initial conditions

The next step in reducing cosmic variance was developed by Angulo & Pontzen (2016). In their approach, the amplitudes of the initial conditions are fixed to the expected mean value by setting

\[
A(k) = \sqrt{\frac{VP_{\text{target}}(|k|)}{(2\pi)^3}}.
\]

The simulations that start from this initial condition are called Fixed simulations. A further reduction of the cosmic variance can be achieved by combining this technique with the paired simulation method described in Sect. 2.1 (Angulo & Pontzen 2016). These paired-and-fixed (PF) simulations can reduce the variance of the power spectrum by factors as large as \( 10^6 \). A detailed description of the statistical properties of the paired and PF simulations can be found in Villaescusa-Navarro et al. (2018). Although the PF method is very effective in reducing the variance of the power spectrum, the halo mass function, and the bispectrum, Klypin et al. (2020) found that the covariances are significantly suppressed, in contrast to the series of traditional simulations with normal Gaussian initial conditions. The main source of this bias is the fact that the initial conditions are missing the random amplitude fluctuations. As this is a well known and very effective method for reducing the cosmic variance, we use the PF simulations as a reference when we compare the different power spectra in this paper.

2.3. Matched pairs

An independent approach was used by Harnois-Déraps et al. (2019) to reduce cosmic variance in cosmic-shear covariance calculations. These authors generated a large ensemble of independent initial conditions and chose two from the set that had an average that was closest to \( P_{\text{target}}(k) \). This pair of initial conditions was used as the starting point for a pair of simulations. Unlike the Paired simulations described in Sect. 2.2, the effects of phase correlations are not canceled out during the averaging of these simulations, as the \( \phi(k) \) phases are completely independent in these pairs.

2.4. Paired-and-mean initial conditions

The first new method we propose in this article is the paired-and-mean (PM) initial conditions. According to this method, the
initial $\delta(k)$ field is generated in the traditional way in the first step, but the initial amplitudes are scaled by

$$A_{\text{PM}}(k) = A(k) \cdot \sqrt{\frac{VP_{\text{target}}(k)}{(2\pi)^3 \langle A(k')^2 \rangle_{k' = k}},}$$  \hspace{1cm} (7)$$

where $A_{\text{PM}}(k)$ are the new PM amplitudes used in the initial condition, and $A(k)$ are the amplitudes generated with the traditional technique. This transformation guarantees that the initial power spectrum is equal to $P_{\text{target}}(k)$, while there are still fluctuations in the amplitudes. We propose that these simulations be run in inverted pairs in order to reduce the effect of phase correlations.

In practice, PM initial conditions can be generated using traditional initial-condition generators such as NGen_IC (Springel 2015) or 2LPTic (Crocce et al. 2012, 2006) using the following target power spectrum:

$$P_{\text{PM}}(k) = \frac{P_{\text{target}}(k)^2}{P_{\text{IC}}(k)},$$  \hspace{1cm} (8)$$

where $P_{\text{IC}}(k)$ is the power spectrum of a traditional initial condition generated with the same code, parameters, and random seed. As the Zel’dovich or 2LPT approximations also have an effect on the final power spectrum (Crocce et al. 2006), $P_{\text{IC}}(k)$ should be calculated from the generated particle distribution.

### 2.5. Complementary initial conditions

Our aim is to use pairs of simulations as part of this new technique, where the initial condition of the first simulation is generated in the traditional way, while the second has modified initial amplitudes and phases. We refer to the former as the “original” simulation, and the latter as the “complementary” simulation. We choose the initial $A_{\text{C}}(k)$ values in the complementary initial condition so that the average of the two initial $P(k)$ matches the $P_{\text{target}}(k)$ power spectrum. This constraint can be written as

$$\left\langle A_{\text{C}}(k')^2 \right\rangle_{k' = k} = \frac{2V}{(2\pi)^3} P_{\text{target}}(k) = \left\langle A(k')^2 \right\rangle_{k' = k},$$  \hspace{1cm} (9)$$

where $A_{\text{C}}(k)$ is the amplitude in the complementary initial condition at $k$ wavenumber. This can only be satisfied for the $k$ wavenumbers where

$$2 \cdot P_{\text{IC}}(k) < P_{\text{target}}(k),$$  \hspace{1cm} (10)$$

because $\langle A(k')^2 \rangle_{k' = k}$ must be a positive. Although this cannot be done for the entire $|k| = k$ surface, this technique can lead to a significant reduction in the variance of the power spectrum by compensating amplitudes that satisfy this criterion. There are an infinite number of $A_{\text{C}}(k)$ fields that satisfy Eq. (9) when Eq. (10) is true. In this study, we chose to generate these fields with the following formula:

$$A_{\text{C}}(k) \mid_{2 \cdot P_{\text{IC}}(k) < P_{\text{target}}(k)} = A(k) \cdot \sqrt{\frac{VP_{\text{target}}(k)}{(2\pi)^3 \langle A(k')^2 \rangle_{k' = k}} \left(2 - \frac{(2\pi)^3 \langle A(k')^2 \rangle_{k' = k}}{VP_{\text{target}}(k)}\right)}.$$  \hspace{1cm} (11)$$

For the modes that cannot be compensated, instead of setting them at zero, we chose

$$A_{\text{C}}(k) \mid_{2 \cdot P_{\text{IC}}(k) \geq P_{\text{target}}(k)} = A(k) \cdot \sqrt{\frac{VP_{\text{target}}(k)}{(2\pi)^3 \langle A(k')^2 \rangle_{k' = k}} \left(2 - \frac{(2\pi)^3 \langle A(k')^2 \rangle_{k' = k}}{VP_{\text{target}}(k)}\right)}$$

in order to avoid introducing any strong unwanted beat-coupling effects (Hamilton et al. 2006). The advantage of this choice is that the complementary initial condition can be generated with the usual codes simply by setting the target power spectrum to

$$P_{\text{input,C}}(k) \mid_{2 \cdot P_{\text{IC}}(k) < P_{\text{target}}(k)} = \frac{2V}{(2\pi)^3} P_{\text{target}}(k) = P_{\text{target}}(k)$$

$$= P_{\text{input,PM}}(k) \cdot \left(2 - \frac{P_{\text{IC}}(k)}{P_{\text{target}}(k)}\right),$$  \hspace{1cm} (13)$$

$$P_{\text{input,C}}(k) \mid_{2 \cdot P_{\text{IC}}(k) \geq P_{\text{target}}(k)} = P_{\text{input,PM}}(k) \cdot \left(\frac{P_{\text{target}}(k)}{P_{\text{IC}}(k)}\right),$$  \hspace{1cm} (14)$$

with the same random seed and parameters, but shifted phases. With this new method, it is possible to provide complementary simulations for existing simulations. By averaging the results of these pairs, a significant reduction in the cosmic variance can be expected, which is similar to the results of the PF simulations. We call these pairs original-complementary (O–C) pairs. It is possible to generate initial conditions without a phase shift, and these are called original-complementary amplitude (O–CA) pairs.

### 3. Comparing the different simulation methods

To test the effects of the new initial-condition-generation techniques, we ran four sets of cosmological N-body simulations. A summary of the simulations is provided in Table 1. All simulations were run with GADGET-4 (Springel et al. 2021), and two had the Planck 2018 cosmological parameters (Planck Collaboration VI 2020), while two used the parameters of the Millennium run (Springel et al. 2005). In the $L_{\text{box}} = 260.0 \text{ Mpc} h^{-1}$ and $L_{\text{box}} = 800.0 \text{ Mpc} h^{-1}$ sets of simulations, there were a few modes where Eq. (10) was not satisfied: only one mode in the $L_{\text{box}} = 260.0 \text{ Mpc} h^{-1}$ set and two modes in the $L_{\text{box}} = 800.0 \text{ Mpc} h^{-1}$ series. To compare the power spectra of the different simulation techniques, we used the PF power spectrum as a reference in the $L_{\text{box}} = 260.0 \text{ Mpc} h^{-1}$ and $L_{\text{box}} = 800.0 \text{ Mpc} h^{-1}$ simulation sets, and calculated the reduced chi-squared statistic as a function of the redshift, plotted in the top panel of Fig. 1. The same quantity was calculated

$$\chi^2(z) = \frac{1}{v} \sum_i \frac{(P_M(k_i, z) - P_{\text{PF}}(k_i, z))^2}{\sigma^2(k_i, z)},$$  \hspace{1cm} (15)$$

with the same random seed and parameters, but shifted phases. With this new method, it is possible to provide complementary simulations for existing simulations. By averaging the results of these pairs, a significant reduction in the cosmic variance can be expected, which is similar to the results of the PF simulations. We call these pairs original-complementary (O–C) pairs. It is possible to generate initial conditions without a phase shift, and these are called original-complementary amplitude (O–CA) pairs.
for the $L = 800$ Mpc $h^{-1}$ simulation set, and this can be seen in the bottom panel of the same figure. Although the original simulations in both sets were close to $\chi^2_{O} = 1$ at early times, this value increased quickly as the matter field evolved. The power spectra from the P and O–CA simulations perform significantly better than the original simulations, but their $\chi^2$ values are still above 1 at a redshift of $z = 0$. As expected, the PM and O–C simulations perform best during these tests: the $\chi^2_{PM}$ and $\chi^2_{O-C}$ values remain below 1 throughout the entire simulated redshift range. An additional visualization of the effect of the new method on the power spectrum is shown in Fig. 2, which shows the ratio of the O–C and PF power spectra in the $L = 800$ Mpc $h^{-1}$ simulations with 1% and 1σ deviations. The complementary simulation method significantly reduces the cosmic variance of the power spectrum at all simulated redshifts, and has a sub-percent accuracy for all compensated modes, even at the largest resolved scales.

The final simulation set contains a single O–C pair. Each simulation in the pair has $L = 500$ Mpc $h^{-1}$ linear size with 10 billion particles and requires 0.74 million CPU hours to run. We used this set in Sect. 4 to measure the differences between two independent O–C realizations of the same cosmology. We call this simulation pair “NewMillennium”.

### 4. Complementary Millennium-run

To further demonstrate the effectiveness of the O–C pair method in reducing cosmic variance, we planned to run the complementary pair of the original Millennium run. This dark matter only TreePM simulation contains $2160^3$ particles in a periodic box with a linear size of $L_{box} = 500$ Mpc $h^{-1}$, and follows the evolution of cosmic structures from redshift $z = 127$ to $z = 0$ with more than 10 000 time steps. The cosmological parameters are the following: $\Omega_m = 0.25$, $\Omega_b = 0.045$, $h = 0.73$, $\Omega_\Lambda = 0.75$, $n = 1$, and $\sigma_8 = 0.9$. As using the same Lean-Gadget-2 code

**Table 1. Parameters of simulations.**

| Simulation sets | L260_N20M | L800_N531M | MillenniumG4 | NewMillennium |
|-----------------|-----------|------------|--------------|--------------|
| IC generator    | 2LPTic    | N-GenIC    | L-GenIC      | L-GenIC      |
| Initial redshift| 31        | 127        | 127          | 127          |
| Final redshift  | 0         | 0          | 0            | 1            |
| $L_{box}$ [Mpc $h^{-1}$] | 260 | 800  | 800          | 800          |
| $N_{part}$      | 270$^3$   | 810$^3$    | 2160$^3$     | 2160$^3$     |
| Cosmology       | Planck2018| Planck2018 | Millennium   | Millennium   |
| Initial conditions | O, P, PF, PM, CA, C | O, P, PF, C | O, C | O, C |

**Notes.** The abbreviations of the different initial condition types are: O – original; P – phase shifted by $\pi$; PF – paired-and-fixed; PM – paired-and-mean (PM); CA – complementary amplitudes without phase shift; C – complementary.
Fig. 3. Dark matter density field of the Millennium run (left) and its complementary pair (right) at $z = 0$ redshift. The bottom row is a zoom onto the region outlined by the red boxes in the figures in the top row. As the $\varphi(k)$ phases are shifted by $\pi$ in the initial condition, a region that collapses to a halo in the original Millennium run tends to expand into a void in the corresponding complementary pair and vice versa, which is similar to the paired simulation method.

(Springel et al. 2005; Springel 2005) that was used to run the original simulation is not possible on modern systems, we chose to use the more recent Gadget-4 for this task. To estimate the effects of using a different simulation code, we re-simulated the small volume version of the Millennium run called “Milli-Millennium” with Gadget-4. We find a percent-level discrepancy in the power spectrum at the final state between the results of the LeanGadget-2 and Gadget-4 simulations. We conclude that these differences emerge from a restart of the original simulation, and from some unknown differences in the default setups for the transition between the particle mesh and tree force calculations. As we were unable to minimize the effects emerging from using different versions of the same code, we re-simulated the original Millennium run with Gadget-4. We used the resources of the Texas Advanced Computing Center (TACC) to run the Millennium simulation and its complementary pair.

During the initial condition creation of the complementary pair, we were able to compensate all modes of the original initial condition, because all $k_i$ bins satisfy Eq. (10). The mass
distributions of both simulations are plotted in Fig. 3. The dimensionless power spectrum of both simulation and the estimated cosmic average power spectrum can be seen in Fig. 4. To further show that the new simulation method produces consistent results, we compared the power spectrum of the Millennium complementary pair with the average power spectrum of the NewMillennium simulation set in Fig. 5. Using the complementary pair, we estimated the bias of the power spectrum of the original Millennium run due to the initial sampling variance and mode coupling by calculating the weighted mean power spectrum ratio

\[ W = \sum_{k_{\text{min}} \leq k \leq k_{\text{max}}} \Delta N_{i,j} P_{\text{mill}}(k)/P_{\text{OC}}(k), \tag{17} \]

where \( P_{\text{mill}}(k) \) is the Millennium power spectrum, and \( P_{\text{OC}}(k) \) is the averaged O-C power spectrum. We find that the original Millennium run, on average, underestimated the power spectrum by 0.997% for \( k \leq 1.0 \text{ h Mpc}^{-1} \) scales and by 0.0881% for \( 50 \text{ h Mpc}^{-1} \geq k \geq 1.0 \text{ h Mpc}^{-1} \) wavenumbers at the \( z = 0 \) redshift.

While by itself the original simulation is not large enough to resolve the baryonic wiggles in the power spectrum due to the sample variance at the low-\( k \) modes, the average of the two new complementary simulations is able to effectively follow the evolution of these baryon acoustic oscillation (BAO) features, as it can be seen in Fig. 6 with the linear and estimated nonlinear spectrum calculated by Code for Anisotropies in the Microwave Background (CAMB: Lewis et al. 2000). We calculated similar results from the original simulation, but this was only achievable by rescaling the linear initial power spectrum by the estimated scale-dependent nonlinear growth function of the Millennium run.

The new Millennium run and its complementary pair are available to the public on the SciServer platform hosted by the Institute for Data Intensive Engineering and Science at the Johns Hopkins University (Taghizadeh-Popp et al. 2020). We made all particle data, halo (Davis et al. 1985) and subhalo (Springel et al. 2001) catalogs, and power spectra available here.

5. Conclusions

1. We reviewed the traditional initial condition generation method, and propose two new techniques for pairs of simulations with reduced cosmic variance: the paired-and-mean and the complementary simulation method.
2. We compared the power spectrum calculated from the new methods with the paired-and-fixed method, and show that the paired-and-mean and original-complementary pairs produce similar results to the paired-and-fixed method.
3. A covariance estimation is left to future work, but based on the fact that the initial amplitudes are not fixed in the new methods, we can expect more precise results than those achievable with the paired-and-fixed method.
4. We show that complementary pairs for existing simulations can be generated, and the sample-variance errors of the original run can be estimated with this pair due to the fact that structures are evolving in the opposite way in the new simulation.
5. To demonstrate the effectiveness of the original-complementary method, we generated the complementary pair of the Millennium-run. Using this new simulation, we show that the original simulation underestimates the power spectrum at all scales at the sub-percent level. The average power spectrum of these two simulations is able to directly resolve the BAO features of the power spectrum.

In this paper, we demonstrated that our new methods can effectively reduce the cosmic variance in \( N \)-body simulations. These new methods will be useful in making predictions for future surveys, testing cosmological models, and estimating errors of independent cosmological simulations.

https://www.sciserver.org/
Fig. 5. Estimated standard deviation (top) and bias (bottom) of the Millennium simulation and its complementary pair. We used the NewMillennium simulation pair as a reference. For this plot, we used 1100 wave number bins with $\Delta k = 9 \times 10^{-3} h \text{Mpc}^{-1}$ bin size in $0.0125 h \text{Mpc}^{-1} < k < 0.10 h \text{Mpc}^{-1}$ range. While individually the two simulations show significant variance in this wave number range and binning, the average of the pair matches the independent New Millennium average power spectrum with sub-percent accuracy.

Fig. 6. Power spectra of the dark matter density field in the Millennium run and its complementary pair at the BAO scales. All data points have been divided by a linear dark matter-only power spectrum. While the original Millennium simulation cannot resolve these scales by itself due to sample variance, the average spectrum of the original and complementary run can efficiently follow the evolution of the BAO features and matches well at the linear scales with the linear CAMB power spectra.

Acknowledgements. GR’s research was supported by an appointment to the NASA Postdoctoral Program administered by Oak Ridge Associated Universities under contract with NASA. GR and AK were supported by JPL, which is run under contract by California Institute of Technology for NASA. This work was supported by the Ministry of Innovation and Technology NRDI Office grants OTKA NN 129148 and the MILAB Artificial Intelligence National Laboratory Program. IS acknowledges support from the National Science Foundation (NSF) award 1616974. The authors thank Volker Springel for significant help with the original and complementary Millennium initial conditions, and for assistance in the comparison of the different Gadget versions. We thank Gerard Lemson for making it possible to store the simulation data in the SciServer platform. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC and visualization resources that have contributed to the research results reported within this paper, URL: http://www.tacc.utexas.edu

References

Aligood, B., Blumenthal, G., & Primack, J. R. 2001, ArXiv e-prints [arXiv:astro-ph/0109463]
Angulo, R. E., & Pontzen, A. 2016, MNRAS, 462, L1
Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, MNRAS, 426, 2046
Baugh, C. M., & Efstathiou, G. 1994, MNRAS, 267, 323
Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150
Casarini, L., Bonometto, S. A., Borgani, S., et al. 2012, A&A, 542, A126
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:1612.05569]
Crocce, M., Pueblas, S., & Scoccimarro, R. 2006, MNRAS, 373, 369

Crocce, M., Pueblas, S., & Scoccimarro, R. 2012, Astrophysics Source Code Library [record ascl:1201.005]
Davies, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
DESI Collaboration (Aghanim, A., et al.) 2016, ArXiv e-prints [arXiv:1611.00936]
Doré, O., Bock, J., Ashby, M., et al. 2014, ArXiv e-prints [arXiv:1412.4872]
Driver, S. P., & Robotham, A. S. G. 2010, MNRAS, 407, 2131
Falck, B., Wang, J., Jenkins, A., et al. 2021, MNRAS, 506, 2659
Green, J., Schechter, P., Baltay, C., et al. 2012, ArXiv e-prints [arXiv:1208.4012]
Hamilton, A. J. S., Rimes, C. D., & Scoccimarro, R. 2006, MNRAS, 371, 1188
Harnois-Dérap, J., Giblin, B., & Joachimi, B. 2019, A&A, 631, A160
Heitmann, K., Finkel, H., Pope, A., et al. 2019, ApJS, 245, 16
Hockney, R. W., & Eastwood, J. W. 1988, Computer Simulation Using Particles (Bristol: Hilger)
Jenkins, A. 2010, MNRAS, 403, 1859
Jenkins, A., Frenk, C. S., White, S. D. M., et al. 2001, MNRAS, 321, 372
Klypin, A., & Holtzman, J. 1997, ArXiv e-prints [arXiv:astro-ph/9712217]
Klypin, A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102
Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340
Klypin, A., Prada, F., & Byun, J. 2020, MNRAS, 496, 3862
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
LSST Science Collaboration (Abell, P. A., et al.) 2009, ArXiv e-prints [arXiv:0912.0281]
Maddox, S. J., Sutherland, W. J., Efstathiou, G., & Loveday, J. 1990, MNRAS, 243, 692
Maksimova, N. A., Garrison, L. H., Eisenstein, D. J., et al. 2021, MNRAS, 508, 4017
McBride, C., Berlind, A., Scoccimarro, R., et al. 2009, Am. Astron. Soc. Meet. Abstr., 213, 425.06
Moster, B. P., Somerville, R. S., Newman, J. A., & Rix, H.-W. 2011, ApJ, 731, 113
