Learning to predict the cosmological structure formation

Siyu Heab,c,1, Yin Liab,d,f, Yu Fengde, Shirley Hoab,c,d,e,1, Siamak Ravanbakhshg, Wei Chen1, and Barnabás Pócozh

Abstract

Matter evolved under the influence of gravity from minuscule density fluctuations. Nonperturbative structure formed hierarchically over all scales and developed non-Gaussian features in the Universe, known as the cosmic web. To fully understand the structure formation of the Universe is one of the holy grails of modern astrophysics. Astrophysicists survey large volumes of the Universe and use a large ensemble of computer simulations to compare with the observed data to extract the full information of our own Universe. However, to evolve billions of particles over billions of years, even with the simplest physics, is a daunting task.

We build a deep neural network, the Deep Density Displacement Model (D3M), which learns from a set of prerin simulation predictions to predict the nonlinear large-scale structure of the Universe with the Zel’dovich Approximation (ZA), an analytical approximation based on perturbation theory, as the input. Our extensive analysis demonstrates that D3M outperforms the second-order perturbation theory (ZP2), the commonly used fast-approximate simulation method, in predicting cosmic structure in the nonlinear regime. We also show that D3M is able to accurately extrapolate far beyond its training data and predict structure formation for significantly different cosmological parameters. Our study proves that deep learning is a practical and accurate alternative to approximate 3D simulations of the gravitational structure formation of the Universe.

Significance

To understand the evolution of the Universe requires a concerted effort of accurate observation of the sky and fast prediction of structures in the Universe. N-body simulation is an effective approach to predicting structure formation of the Universe, though computationally expensive. Here, we build a deep neural network to predict structure formation of the Universe. It outperforms the traditional fast-analytical approximation and accurately extrapolates far beyond its training data. Our study proves that deep learning is an accurate alternative to the traditional way of generating approximate cosmological simulations. Our study shows that one can use deep learning to generate complex 3D simulations in cosmology. This suggests that deep learning can provide a powerful alternative to traditional numerical simulations in cosmology.

Author contributions: S. He, Y.L., S. Ho, and B.P. designed research; S. He, Y.L., Y.F., and S. Ho performed research; S. He, Y.L., S.R., and B.P. contributed new reagents/analytic tools; S. He, Y.L., Y.F., and W.C. analyzed data; and S. He, Y.L., Y.F., S. Ho, S.R., and W.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The source code of our implementation is available at https://github.com/rainwoodman/fastpm.y

Author affiliations: 1Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213; 2McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213; 3Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010; 4Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720; 5Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; 6Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo Institutes for Advanced Study, The University of Tokyo, Chiba 277-8583, Japan; 7Computer Science Department, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; and 8Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213

www.pnas.org/cgi/doi/10.1073/pnas.1821458116

PNAS Latest Articles | 1 of 8
The outcome of a typical N-body simulation depends on both the initial conditions and on cosmological parameters which affect the evolution equations. A striking discovery is that the Deep Density Displacement Model (D^3M), trained by using a single set of cosmological parameters, generalizes to new sets of significantly different parameters, minimizing the need for training data on a diverse range of cosmological parameters.

**Setup**

We build a deep neural network, D^3M, with similar input and output of an N-body simulation. The input to our D^3M is the displacement field from the Zel’dovich Approximation (ZA) (55). A displacement vector is the difference of a particle position at a special location or direction. This allows one to drop the dependence on a particular location or direction. This allows one to drop the dependence on special location or direction. This allows one to drop the dependence on a special location or direction. This allows one to drop the dependence on a special location or direction.

As suggested by Fig. 2, the denser regions seem to have a higher error for all methods—that is, more nonlinearity in structure formation creates larger errors for both D^3M and 2LPT. The dependence of error on scale is computed with two- and three-point correlation analysis.

**Results and Analysis**

Fig. 1 shows the displacement vector field as predicted by D^3M (Left) and the associated point-cloud representation of the structure formation (Right). It is possible to identify structures such as clusters, filaments, and voids in this point-cloud representation. We proceed to compare the accuracy of D^3M and 2LPT compared with ground truth.

**Point-Wise Comparison.** Let $\Psi \in \mathbb{R}^{d \times d \times 3}$ denote the displacement field, where $d$ is the number of spatial-resolution elements in each dimension ($d = 32$). A natural measure of error is the relative error $|\Psi - \Psi_t|/|\Psi_t|$, where $\Psi_t$ is the true displacement field (FastPM), and $\Psi$ is the prediction from 2LPT or D^3M. Fig. 2 compares this error for different approximations in a 2D slice of a single simulation. We observe that D^3M predictions are very close to the ground truth, with a maximum relative error of 1.10 over all 1,000 simulations. For 2LPT, this number is significantly higher at 4.23. In average, the result of D^3M comes with a 2.8% relative error, while for 2LPT, it equals 9.3%.

**Two-Point Correlation Comparison.** As suggested by Fig. 2, the denser regions seem to have a higher error for all methods—that is, more nonlinearity in structure formation creates larger errors for both D^3M and 2LPT. The dependence of error on scale is computed with two- and three-point correlation analysis.

Cosmologists often use compressed summary statistics of the density field in their studies. The most widely used of these statistics are the two-point correlation function (2PCF) $\xi(r)$ and its Fourier transform, the power spectrum $P_{\delta\delta}(k)$:

$$\xi(r) = \langle \delta_A(r')\delta_B(r + r') \rangle,$$

$$P_{\delta\delta}(k) = \int d^3r \xi(r)e^{ikr},$$

where the ensemble average $\langle \rangle$ is taken over all possible realizations of the Universe. Our Universe is observed to be both homogeneous and isotropic on large scales—i.e., without any special location or direction. This allows one to drop the dependencies on $r$ and on the direction of $r$, leaving only the amplitude $|\xi|$ in the final definition of $\xi(r)$. In the second equation, $P_{\delta\delta}(k)$ is simply the Fourier transform of $\xi(r)$ and captures the dispersion of the plane-wave amplitudes at different scales in the Fourier space. $k$ is the 3D wavevector of the plane wave, and its amplitude $k$ (the wavenumber) is related to the wavelength $\lambda$ by $k = 2\pi/\lambda$. Due to isotropy of the Universe, we drop the vector form of $r$ and $k$.

![Fig. 1. The displacement vector field (Left) and the resulting density field (Right) produced by D^3M. The vectors in Left are uniformly scaled down for better visualization.](image)

---

He et al.

2 of 8 | www.pnas.org/cgi/doi/10.1073/pnas.1821458116

---
Because FastPM, 2LPT, and D^3M take the displacement field as input and output, we also study the two-point statistics for the displacement field. The displacement power spectrum is defined as:

\[ P_{\psi\psi}(k) = \langle \psi_\alpha(k)\psi_\beta^*(k) \rangle + \langle \psi_\alpha(k)\psi_\beta^*(k) \rangle + \langle \psi_\alpha(k)\psi_\beta^*(k) \rangle. \]  

We focus on the Fourier-space representation of the two-point correlation. Because the matter and the displacement power spectrum take the same form, in what follows, we drop the subscript for matter and displacement field and use \( P(k) \) to stand for both matter and displacement power spectrum. We use the transfer function \( T(k) \) and the correlation coefficient \( r(k) \) as metrics to quantify the model performance against the ground truth (FastPM) in the two-point correlation. We define the transfer function \( T(k) \) as the square root of the ratio of two power spectra,

\[ T(k) = \sqrt{\frac{P_{\text{pred}}(k)}{P_{\text{true}}(k)}}, \]

where \( P_{\text{pred}}(k) \) is the density or displacement power spectrum as predicted by 2LPT or D^3M, and, analogously, \( P_{\text{true}}(k) \) is the ground truth predicted by FastPM. The correlation coefficient \( r(k) \) is a form of normalized cross-power spectrum,

\[ r(k) = \frac{P_{\text{pred,\text{true}}}(k)}{\sqrt{P_{\text{pred}}(k)P_{\text{true}}(k)}}, \]

where \( P_{\text{pred,\text{true}}}(k) \) is the cross-power spectrum between 2LPT or D^3M predictions and the ground-truth (FastPM) simulation result. The transfer function captures the discrepancy between amplitudes, while the correlation coefficient can indicate the discrepancy between phases as functions of scales. For a perfectly accurate prediction, \( T(k) \) and \( r(k) \) are both 1. In particular, \( 1 - r^2 \) describes stochasticity, the fraction of the variance in the prediction that cannot be explained by the true model.

Fig. 3A shows the average power spectrum, transfer function \( T(k) \), and stochasticity \( 1 - r^2(k) \) of the displacement field and the density field over 1,000 simulations. The transfer function of density from 2LPT predictions is 2% smaller than that of FastPM on large scales (\( k \approx 0.05 \text{ hMpc}^{-1} \)). This is expected since 2LPT performs accurately on very large scales (\( k < 0.01 \text{ hMpc}^{-1} \)). The displacement transfer function of 2LPT increases \( >1 \) at \( k \approx 0.35 \text{ hMpc}^{-1} \) and then drops sharply. The increase of the 2LPT displacement transfer function is because 2LPT overestimates the displacement power at small scales (e.g., ref. 65). There is a sharp drop of power near the voxel scale because smoothing over voxel scales in our predictions automatically erases power at scales smaller than the voxel size.

Now, we turn to the D^3M predictions: Both the density and displacement transfer functions of the D^3M differ from 1 by a mere 0.4% at scale \( k \lesssim 0.4 \text{ hMpc}^{-1} \), and this discrepancy only increases to 2% and 4% for density field and displacement field, respectively, as \( k \) increases to the Nyquist frequency at \( 0.7 \text{ hMpc}^{-1} \). The stochasticity hovers at \( \sim 10^{-3} \) and \( 10^{-2} \) for most scales. In other words, for both the density and displacement fields, the correlation coefficient between the D^3M predictions and FastPM simulations, all the way down to small scales of \( k = 0.7 \text{ hMpc}^{-1} \), is >90%. The transfer function and correlation coefficient of the D^3M predictions show that it can reproduce the structure formation of the Universe from large to semilinear scales. D^3M significantly outperforms our benchmark model 2LPT in the two-point function analysis. D^3M only starts to deviate from the ground truth at fairly small scales. This is not surprising, as the deeply nonlinear evolution at these scales is more difficult to simulate accurately and appears to be intractable by current analytical theories (66).

Three-Point Correlation Comparison. The three-point correlation function (3PCF) expresses the correlation of the field of interest among three locations in the configuration space, which is equivalently defined as bispectrum in Fourier space. Here, we concentrate on the 3PCF for computational convenience:

\[ \zeta(r_1, r_2, \theta) = \langle \delta(x)\delta(x + r_1)\delta(x + r_2) \rangle, \]

where \( r_1 = |r_1| \) and \( r_2 = |r_2| \). Translation invariance guarantees that \( \zeta \) is independent of \( x \). Rotational symmetry further eliminates all direction dependence except dependence on \( \theta \), the angle between \( r_1 \) and \( r_2 \). The multipole moments of \( \zeta(r_1, r_2, \theta) \),
\[ \zeta(r_1, r_2) = (2\ell + 1) \int \, d\theta P_{\ell}(\cos \theta) \zeta(r_1, r_2, \theta), \] where \( P_{\ell}(\cos \theta) \) is the Legendre polynomial of degree \( \ell \), can be efficiently estimated with pair counting (67). While the input (computed by ZA) do not contain significant correlations beyond the second order (power spectrum level), we expect \( D^M \) to generate densities with a 3PCF that mimics that of ground truth.

We compare the 3PCF calculated from FastPM, 2LPT, and \( D^M \) by analyzing the 3PCF through its multipole moments \( \zeta(\ell, r_1, r_2) \). Fig. 3B shows the ratio of the binned multipole coefficients of the two 3PCFs for several triangle configurations, \( \bar{\zeta}(\ell, r_1, r_2)_{\text{pred}} / \bar{\zeta}(\ell, r_1, r_2)_{\text{true}} \), where \( \bar{\zeta}(\ell, r_1, r_2)_{\text{pred}} \) can be the 3PCF for \( D^M \) or 2LPT and \( \bar{\zeta}(\ell, r_1, r_2)_{\text{true}} \) is the 3PCF for FastPM. We used 10 radial bins with \( \Delta r = 5 \, h^{-1} \text{Mpc} \). The results are averaged over 10 test simulations, and the error bars are the SD. The ratio shows that the 3PCF of \( D^M \) is closer to FastPM than 2LPT, with smaller error bars. To further quantify our comparison, we calculate the relative 3PCF residual defined by

\[ \text{3PCF relative residual} = \frac{1}{9 \times N_r} \sum_{\ell=0}^{8} \sum_{r_1, r_2} \frac{|\zeta(\ell, r_1, r_2)_{\text{pred}} - \zeta(\ell, r_1, r_2)_{\text{true}}|}{|\zeta(\ell, r_1, r_2)_{\text{true}}|}. \tag{8} \]

where \( N_r \) is the number of \((r_1, r_2)\) bins. The mean relative 3PCF residual of the \( D^M \) and 2LPT predictions compared with FastPM are 0.79% and 7.82%, respectively. The \( D^M \) accuracy on 3PCF is also an order of magnitude better than 2LPT, which indicates that the \( D^M \) is far better at capturing the non-Gaussian structure parameters.

**Generalizing to New Cosmological Parameters**

So far, we train our model using a “single” choice of cosmological parameters \( A_s = 2.142 \times 10^{-9} \) (hereafter \( A_0 = 2.142 \times 10^{-9} \)) and \( \Omega_m = 0.3089 \) (68). \( A_s \) is the primordial amplitude of the scalar perturbation from cosmic inflation, and \( \Omega_m \) is the fraction of the total energy density that is matter at the present time, and we will call it matter density parameter for short. The true exact value of these parameters is unknown, and different choices of these parameters change the large-scale structure of the Universe; Fig. 4.

Here, we report an interesting observation: The \( D^M \) trained on a single set of parameters in conjunction with ZA (which depends on \( A_s \) and \( \Omega_m \)) as input can predict the structure formation for widely different choices of \( A_s \) and \( \Omega_m \). From a computational point of view, this suggests a possibility of producing simulations for a diverse range of parameters, with minimal training data.

**Varying Primordial Amplitude of Scalar Perturbations \( A_s \)**

After training the \( D^M \) using \( A_s = A_0 \), we change \( A_s \) in the input of our test set by nearly one order of magnitude: \( A_s = 1.8A_0 \) and \( A_s = 0.2A_0 \). Again, we use 1,000 simulations for analysis of each test case. The average relative displacement error of \( D^M \) remains <4% per voxel (compared with <3% when train and test data have the same parameters). This is still well below the error for 2LPT, which has relative errors of 15.5% and 6.3% for larger and smaller values of \( A_s \), respectively.

Fig. 5A shows the transfer function and correlation coefficient for both \( D^M \) and 2LPT. The \( D^M \) performs much better than 2LPT for \( A_s = 1.8A_0 \). For small \( A_s = 0.2A_0 \), 2LPT does a better job than \( D^M \) predicting the density transfer function and correlation coefficient at the largest scales; otherwise, \( D^M \) predictions are more accurate than 2LPT at scales larger than \( k < 0.08 \, h\, \text{Mpc}^{-1} \). We observe a similar trend with 3PCF analysis: The 3PCFs of \( D^M \) predictions are notably better than 2LPT ones for larger \( A_s \), compared with smaller \( A_s \), where it is only slightly better. These results confirm our expectation that increasing \( A_s \) increases the non-linearity of the structure-formation process. While 2LPT can predict fairly well in linear regimes, compared with \( D^M \), its performance deteriorates with increased nonlinearity. It is interesting to note that the \( D^M \) prediction maintains its advantage, despite being trained on data from more linear regimes.
Varying Matter Density Parameter $\Omega_m$. We repeat the same experiments, this time changing $\Omega_m$ to 0.5 and 0.1, while the model is trained on $\Omega_m = 0.3089$, which is quite far from both of the test sets. For $\Omega_m = 0.5$, the relative residual displacement errors of the $D^3M$ and 2LPT averaged over 1,000 simulations are 3.8% and 15.2%, and for $\Omega_m = 0.1$, they are 2.5% and 4.3%. Fig. 5 C and D show the two-point statistics for density field predicted by using different values of $\Omega_m$. For $\Omega_m = 0.5$, the results show that the $D^3M$ outperforms 2LPT at all scales, while for smaller $\Omega_m = 0.1$, $D^3M$ outperforms 2LPT on smaller scales ($k > 0.1 \, h \text{Mpc}^{-1}$). As for the 3PCF of simulations with different values of $\Omega_m$, the mean relative 3PCF residual of the $D^3M$ for $\Omega_m = 0.5$ and $\Omega_m = 0.1$ are 1.7% and 1.2%, respectively, and for 2LPT, they are 7.6% and 1.7%, respectively. The $D^3M$ prediction performs better at $\Omega_m = 0.5$ than $\Omega_m = 0.1$. This is again because the Universe is much more nonlinear at $\Omega_m = 0.5$ than $\Omega_m = 0.1$. The $D^3M$ learns more nonlinearity than is encoded in the formalism of 2LPT.

Conclusions

To summarize, our deep model $D^3M$ can accurately predict the large-scale structure of the Universe as represented by FastPM simulations, at all scales, as seen in Table 1. Furthermore, $D^2M$ learns to predict cosmic structure in the nonlinear regime more accurately than our benchmark model 2LPT. Finally, our model generalizes well to test simulations with cosmological parameters ($A_s$ and $\Omega_m$) significantly different from the training set. This suggests that our deep-learning model can potentially be deployed for a range of simulations beyond the parameter space covered by the training data (Table 1). Our results demonstrate that the $D^3M$ successfully learns the nonlinear mapping from first-order perturbation theory to FastPM simulation beyond what higher-order perturbation theories currently achieve.

Looking forward, we expect that replacing FastPM with exact $N$-body simulations would improve the performance of our method. As the complexity of our $D^3M$ model is linear in the

Fig. 4. We show the differences of particle distributions and displacement fields when we change the cosmological parameters $A_s$ and $\Omega_m$. (A) The error bar shows the difference of particle distribution (Upper) and displacement fields (Lower) between $A_s = A_0$ and the two extremes for $A_s = 0.2 \, A_0$ (Center) and $A_s = 1.8 \, A_0$ (Right). (B) A similar comparison showing the difference of the particle distributions (Upper) and displacement fields (Lower) for smaller and larger values of $\Omega_m \in \{0.1, 0.5\}$ with regard to $\Omega_m = 0.3089$, which was used for training. While the difference for the smaller value of $A_s$ ($\Omega_m$) is larger, the displacement for the larger $A_s$ ($\Omega_m$) is more nonlinear. This nonlinearity is due to concentration of mass and makes the prediction more difficult.

Fig. 5. Similar plots as in Fig. 3A, except that we test the two-point statistics when we vary the cosmological parameters without changing the training set (which has different cosmological parameters) or the trained model. We show predictions from $D^3M$ and 2LPT when tested on different $A_s$ (A) and $\Omega_m$ (B). We show the transfer function—i.e., the square root of the ratio of the predicted power spectrum to the ground truth (Upper)—and $1 - r^2$, where $r$ is the correlation coefficient between the predicted fields and the true fields (Lower). The $D^3M$ prediction outperforms 2LPT prediction at all scales except in the largest scales, as the perturbation theory works well in linear regime (large scales).
number of voxels, we expect to be able to further improve our results if we replace the FastPM simulations with high-resolution simulations. Our work suggests that deep learning is a practical and accurate alternative to the traditional way of generating approximate simulations of the structure formation of the Universe.

Materials and Methods

Dataset. The full simulation data consists of 10,000 simulations of boxes with ZA and FastPM as input–output pairs, with an effective volume of 20 (Gpc h)$^3$ ($190 \times 10^9 h^3$), comparable to the volume of a large spectroscopic sky survey like Dark Energy Spectroscopic Instrument or EUCLID. We split the full simulation dataset into 80%, 10%, and 10% for training, validation, and test, respectively. We also generated 1,000 simulations for 2LPT for each set of tested cosmological parameters.

Model and Training. The $D^3M$ adopts the U-Net architecture (54) with 15 convolution or deconvolution layers and $\sim 8.4 \times 10^6$ trainable parameters. Our $D^3M$ generalizes the standard U-Net architecture to work with 3D data (69–71). The details of the architecture are described in the following sections, and a schematic figure of the architecture is shown in SI Appendix, Fig. S2. In the training phase, we use the Adam Optimizer (72) with a learning rate of 0.0001, and first- and second-moment exponential decay rates of $0.9$ and $0.99$, respectively. We use the mean-squared error as the loss function (Loss Function) and L2 regularization with regularization coefficient 0.0001.

Details of the $D^3M$ Architecture. The contracting path follows the typical architecture of a convolution network. It consists of two blocks, each of which consists of two successive convolutions of stride 1 and a downsampling convolution with stride 2. The convolution layers use $3 \times 3 \times 3$ filters. A special feature of the $D^3M$ is the concatenation procedure, where the upsampling layer halves the feature channels and then concatenates them with the corresponding feature channels on the contracting path, doubling the number of feature channels.

The expansive building block then follows a $1 \times 1 \times 1$ convolution without padding, which converts the 64 features to the final 3D displacement field. All convolutions in the network except the last one are followed by a rectified linear unit activation and batch normalization. Padding and Periodic Boundary. It is common to use constant or reflective padding in deep models for image processing. However, these approaches are not suitable for our setting. The physical model we are learning is constructed on a spatial volume with a periodic boundary condition. This is sometimes also referred to as a torus geometry, where the boundaries of the simulation box are topologically connected—that is, $x_{i+L} = x_i$, where $i$ is the index of the spatial location, and $L$ is the periodicity (size of box). Constant or reflective padding strategies break the connection between the physically nearby points separated across the box, which not only loses information but also introduces noise during the convolution, further aggravated with an increased number of layers.

We find that the periodic padding strategy significantly improves the performance and expedites the convergence of our model, comparing to the same network using a constant padding strategy. This is not surprising, as one expects that it is easier to train a model that can explain the data than to train a model that does not.

Loss Function. We train the $D^3M$ to minimize the mean square error on particle displacements

$$L = \frac{1}{N} \sum (\Psi_i - \hat{\Psi}_i)^2,$$  

where $i$ labels the particles and $N$ is the total number of particles. This loss function is proportional to the integrated squared error, and by using a
Fourier transform and Parseval’s theorem, it can be rewritten as
\[ \langle \mathbf{F} \mathbf{w} - \mathbf{F} \mathbf{v}, \mathbf{F}^* \mathbf{d} \rangle = \langle \mathbf{F} \mathbf{w} - \mathbf{F} \mathbf{v}, \mathbf{F}^* \mathbf{d} \rangle = \left\| \mathbf{F} \mathbf{w} - \mathbf{F} \mathbf{v} \right\|^2 = \left\| \mathbf{d} \right\|^2 \langle \mathbf{F}^* \mathbf{F} \mathbf{w} - \mathbf{F}^* \mathbf{F} \mathbf{v}, \mathbf{d} \rangle = \left\| \mathbf{d} \right\|^2 \langle \mathbf{w} - \mathbf{v}, \mathbf{d} \rangle = \left\| \mathbf{d} \right\|^2 \left( \mathbf{w}^t \mathbf{1} - \mathbf{v}^t \mathbf{1} \right) = \left\| \mathbf{d} \right\|^2 \left( \mathbf{w}^t \mathbf{1} - \mathbf{v}^t \mathbf{1} \right), \tag{10} \]

where \( q \) is the Lagrangian space position, and \( k \) its corresponding wavevector. \( T \) is the transfer function defined in Eq. 5, and \( r \) is the correlation coefficient defined in Eq. 6, which characterize the similarity between the predicted and true fields, in amplitude and phase, respectively. Eq. 10 shows that our simple loss function jointly captures both of these measures: As \( T \) and \( r \) approach 1, the loss function approaches 0.

**Data Availability.** The source code of our implementation is available at https://github.com/siyucosmo/ML-Recon. The code to generate the training data is also available at https://github.com/rainwoodman/Fastpm.

**ACKNOWLEDGMENTS.** We thank Angus Beane, Peter Braam, Gabriella Contardo, David Hogg, Laurentine Levesque, Pascal Ripoche, Zack Slepian, and Dan Spergel for useful comments and comments; Angus Beane for comments on the paper; and Nick Carriero for help on Center for Computational Astrophysics (CCA) computing clusters. The work was supported partially by the Simons Foundation. The FastPM simulations were generated on the computer cluster Edison at the National Energy Research Scientific Computing Center, a US Department of Energy Office of Science User Facility operated under Contract DE-AC02-05CH11231. The training of the neural network model was performed on the CCA computing facility and the Carnegie Mellon University Advanced Computing Facility. The open-source software toolkit nbtkidy (73) was used for the clustering analysis. Y.L. was supported by the Berkeley Center for Cosmological Physics and the Kavli Institute for the Physics and Mathematics of the Universe, established by the World Premier International Research Center Initiative of the MEXT, Japan. S. Ho was supported by NASA Grants 15-WFIRST5-0008 and Research Opportunities in Space and Earth Sciences Grant 12-EUCLID12-0004, and the Simons Foundation.


