Hybrid Physical-Neural ODEs for Fast N-body Simulations

Denise Lanzieri [* 1]   **François Lanusse** [* 2]   **Jean-Luc Starck** [2]

Abstract

We present a new scheme to compensate for the small-scales approximations resulting from Particle-Mesh (PM) schemes for cosmological N-body simulations. This kind of simulations are fast and low computational cost realizations of the large scale structures, but lack resolution on small scales. To improve their accuracy, we introduce an additional effective force within the differential equations of the simulation, parameterized by a Fourier-space Neural Network acting on the PM-estimated gravitational potential. We compare the results for the matter power spectrum obtained to the ones obtained by the PGD scheme (Potential Gradient Descent scheme). We notice a similar improvement in term of power spectrum, but we find that our approach outperforms PGD for the cross correlation coefficients, and is more robust to changes in simulation settings (different resolutions, different cosmologies). We provide all codes associated with this paper at this link ⚙.

1. Introduction

N-body simulations have become an essential tool for the cosmological analysis of current and upcoming sky surveys. Because solving the full N-body problem can be very expensive in terms of computational resources and time, Quasi N-body numerical schemes have been proposed to speed up the computational time and create low cost realizations of the large scale structure, e.g. FastPM (Feng et al., 2016) and COLA (Tassev et al., 2013). In this work, we adopt a Particle-Mesh (PM) scheme, which approximates the computation of the gravitational forces between the dark matter particles, by computing the gravitational potential on a 3D mesh. The drawback of this PM scheme is that it approximates the short range interactions between particles and

thus is not able to resolve structures with scales smaller than the mesh resolution. As a result, these simulations typically lack power on small scales and halo profiles are less sharp than their full N-body counterparts.

In this work, we propose to improve the accuracy of PM simulations by augmenting the physical differential equations of a PM N-body with a minimally-parametric neural network component modeling a *residual effective force* compensating for the PM approximations. In order to impose rotational and translational invariance, this neural correction is implemented in the form of a learned isotropic Fourier filter.

To train the neural network through the simulator, we follow the scheme proposed by Chen et al. (2018), i.e. we treat the Ordinary Differential Equations (ODE) solver running the simulation as a black box and compute gradients using the adjoint sensitivity method (Pontriagin, 1964). We train and compare the model to the CAMELS simulations (Villaescusa-Navarro et al., 2021) and test it against several changes of the simulation setting (resolution, volume, cosmological parameters). We also compare our method to PGD approach (Potential Gradient Descent scheme) (Dai et al., 2018).

2. Related work

Several machine learning techniques are available to emulate high cost N-body simulations. In particular, He et al. (2019) proposed a deep neural network to learn the nonlinear mapping from first order perturbation theory linear displacements to the displacement field of FastPM simulations. More recently, Li et al. (2021) have demonstrated that a generative model can be trained to super-resolve the particle displacement field, allowing to enhance the resolution of a low cost approximate N-body. These approaches use particle displacements as inputs and outputs of their modeling, but rely on large Deep Convolutional Networks to learn an effective mapping generating the desired outputs.

FastPM   FastPM (Feng et al., 2016) is a quasi particle-mesh (PM) N-body solver, using a specifically designed leapfrog integration scheme modified such that it agrees with the Zeldovich solution on large scales. In this way, at large scale ($k \rightarrow 0$), the correct linear theory growth can be

---

*Equal contribution [1]Université Paris Cité, Université Paris-Saclay, CEA, CNRS, AIM, F-91191, Gif-sur-Yvette, France [2]Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France; . Correspondence to: Denise Lanzieri <denise.lanzieri@cea.fr>.

produced in a limited number of steps.

Potential Gradient Descent (PGD) Dai et al. (2018) introduces a gradient based method to correct the PM approximation in FastPM and improve the modeling of the matter distribution within halos. PGD models the effect of short range interactions as an additional particle displacement term, moving the particles towards a minimum of the gravitational potential after band-pass filtering:

$$\mathbf{S} = \frac{4\pi G\bar{\rho}\alpha_{PM}}{H_0^2} \nabla \hat{\mathbf{O}}_l(k)\hat{\mathbf{O}}_s(k)\nabla^{-2}\delta, \qquad (1)$$

where $\hat{\mathbf{O}}_l(k) = \exp\left(\frac{k^2}{k_l^2}\right)$ and $\hat{\mathbf{O}}_s(k) = \exp\left(-\frac{k^4}{k_s^4}\right)$ are the low and high pass filter introduced to remove the long range force and to reduce the numerical effects induced by the mesh resolution. PGD introduces 3 nuisance parameters fitted on training simulations: α_{PM}, defining the amplitude of the filter and the long and the short range scale parameters k_l and k_s.

3. Hybrid Physical-Neural ODE

Cosmological simulations implement integration of gravitational evolution starting from the Gaussian initial conditions of the Universe to the observed large scale structures.
In this work, we choose to compute the time integration starting from a system of Ordinary Differential Equations (ODE) and treat the ODE solver as a black box:

$$\begin{cases} \frac{d\mathbf{x}}{da} = \frac{1}{a^3 E(a)}\mathbf{v} \\ \frac{d\mathbf{v}}{da} = \frac{1}{a^2 E(a)}F_\theta(\mathbf{x}, a), \end{cases} \qquad (2)$$

where $\mathbf{x}$ and $\mathbf{v}$ define the position and the velocity of the particle and a is the cosmological scale factor. We have also introduced the parametric function F_θ, hybrid between a physical model and a neural network:

$$F_\theta(\mathbf{x}, a) = \frac{3\Omega_m}{2}\nabla\left[\phi_{PM}(\mathbf{x}) * \mathcal{F}^{-1}(1 + f_\theta(a, |\mathbf{k}|))\right], \qquad (3)$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform, ϕ_{PM} is the gravitational potential estimated by using the cloud-in-cell (CiC) interpolation scheme (Hockney & Eastwood, 2021) and $f_\theta(a, |\mathbf{k}|)$ is the learned neural filter, aimed to model the residual effective force compensating for the PM approximations. We choose to define $f_\theta(a, |\mathbf{k}|)$ as a B-spline function, whose coefficients (knot points) are the output of a neural network of parameters θ. More specifically, we choose B-spline of order 3, defined over 16 knot points. We point out that, since we are modeling true physical processes, there are several symmetries we wish to preserve. Therefore, f_θ is implemented as a Fourier-based isotropic filter in order to preserve the translational and rotational symmetries.

To train our hybrid physical-neural ODE and back-propagate through the ODE solver, we compute the gradient adopting the adjoint sensitivity method (Chen et al., 2018; Pontryagin et al., 1962), consisting in solving a second ODE backwards in time, and treat the ODE solver as a black box. We adopt the following loss function penalizing both the positions of the particles and the overall matter power spectrum at different snapshot times s, compared to a reference full N-body simulation:

$$\mathcal{L} = \sum_s ||\mathbf{x}_s^{Nbody} - \mathbf{x}_s||_2^2 + \lambda||\frac{P_s(k)}{P_s^{Nbody}(k)} - 1||_2^2, \quad (4)$$

where $P_s(k)$ is the power spectrum of the simulation for snapshot s, and λ is left as hyper-parameter balancing the contributions of the two terms. In contrast to He et al. (2019) who train the model minimizing the mean square error on particle displacements, we empirically found that penalizing the overall matter power spectrum leads to a better improvement for the small scale power spectrum, compared to penalizing the particle positions only. One possible reason for this result, may be the reduced number of degrees of freedom in our model compared to the one used by He et al. (2019), i.e. in our training process, the neural network cannot approximate arbitrary non-linear functions, it is restricted by physical constraints and our simple parameterisation. However, we know that different displacement fields could produce identical density fields and then, identical power spectrum. If we penalised the power spectrum term only, in principle we may increase the overall power spectrum on small scales, without however recovering the right positions of the particles. For this reason, the final loss function we use includes both terms of Equation 4. We empirically find $\lambda = 0.1$ provides good results. The model is trained by gradient descent with the Adam optimizer (Kingma & Ba, 2014) and learning rates of 0.01. The code is implemented in the Python package Jax (Bradbury et al., 2018) and we use the Jax-based Dormand-Prince ODE solver.

4. Simulation data

In this work, we rely on the CAMELS dataset (Villaescusa-Navarro et al., 2021), in particular we use the suite of dark-matter only N-body simulations implementing the N-body tree-particle-mesh approach of IllustrisTNG (Nelson et al., 2019). We have 34 snapshots, generated following the evolution of 256^3 dark matter particles in a periodic box of comoving volume equal to 25^3 (h^{-1} Mpc)^3, with initial conditions generated at z=127 using second order Lagrangian perturbation theory (2LPT). The following cosmological parameters are kept fixed in all simulations: $h = 0.6711$, $n_s = 0.9624$, $M_\nu = 0.0$ eV, $w = -1$, $\Omega_k = 0.$, while
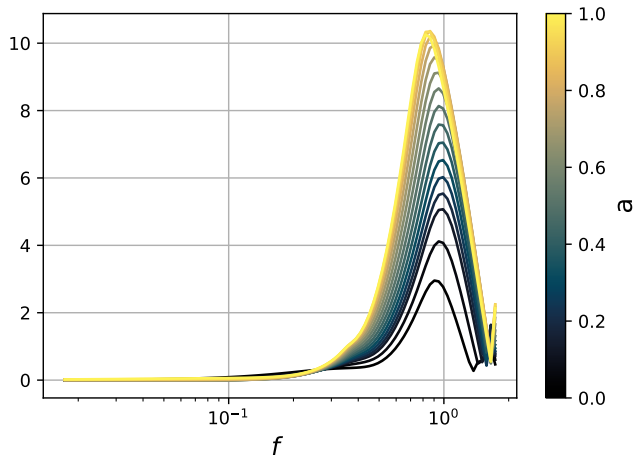
Figure 1. Learned neural filter $f_\theta(a, |\mathbf{k}|)$ as a function of normalized mesh frequency ($f = 1$ corresponds to the Nyquist frequency of the mesh), for different scale factor a.

the values of Ω_m and σ_8 are varied across simulations. We further downsample the CAMELS particle data, down to 64^3 particles, to make the problem more manageable on a single GPU.

5. Results

In this section, we train and evaluate our hybrid neural ODE on CAMELS data, and compare it to results obtained with the PGD approach of Dai et al. (2018). We train the neural ODE using the loss function presented in Equation 4, and the PGD parameters are fitted only for the final snapshot ($a = 1$) by penalizing the weighted power spectrum ratio between the corrected and reference simulations (i.e. Equation 5).

In both cases, we use a *single* CAMELS N-body simulation at the fiducial cosmology of $\Omega_m = 0.3$ and $\sigma_8 = 0.8$ to fit the parameters of the models.

Figure 2 shows the matter overdensity fields obtained with the 2 different correction schemes compared to the CAMELS and the standalone PM (i.e. pure PM simulation). As we will quantify below, we can already see that the pure PM simulation (second panel from the left) is smoother than the reference CAMELS simulation, and that the neural correction (third panel from the left) recovers most of the missing information. By comparison, the PGD correction does sharpen somewhat some structures in the field, in order to increase the overall power spectrum on small scales, but is less successful at restoring the actual shape of the dark matter structures.

Learned Neural Filter   In Figure 1 we show the trained Fourier-space filter $f_\theta(a, |\mathbf{k}|)$ as a function of normalized

mesh frequency. As expected, the correction only affects small scales, having no effect on the large-scale modes of the simulations for which the PM solver is accurate. An interesting feature to note is that the correction is scale-factor dependent. This indicates that the model is not simply applying a sharpening filter akin to a CiC compensation filter, but that the correction most likely adapts to the density field properties at different cosmological times. One drawback of this observation though is that if the neural correction is dependent on the particular dynamics of the dark matter density field, it may be cosmology-dependent.

Power spectrum comparison at fiducial cosmology   We show on Figure 3 a comparison of the ratio of the power spectrum and the cross-correlation of CAMELS and PM simulations to illustrate the difference between the two correction schemes. As can be seen, while most of the missing power is recovered on small scales in both cases, the improvement of cross correlation coefficients after using our scheme is drastically different.

Robustness to changes in resolution and cosmological parameters   To evaluate the robustness of our correction scheme, we compute the following tests:

1. We investigate the effects due to **varying the cosmological parameters**. We run the PM simulations and compare them against a set of CAMELS, both generated at $\Omega_m = 0.10$ (instead of $\Omega_m = 0.3$ used for the training) and the same initial conditions. The top panel of Figure 4 presents the result of this test. Between PGD and NN correction schemes, the NN correction shows a slight deviation compared to results obtained at the fiducial cosmology on Figure 3. PGD, on the contrary, appears to be extremely sensitive to Ω_m, no longer applying the desired correction.

2. We investigate the effects of **increasing the resolution of the simulations**. We run the PM simulation with 128^3 particles for the same 25 $h^{-1} Mpc$ box size. The center panel of Figure 4 shows that the NN correction out-performs PGD up to $k \sim 7$, after which the method overemphasizes the small scale power.

3. Finally, we test the two schemes at **lower resolution**, increasing the volume of the simulation to 200 $(h^{-1}\mathrm{Mpc})^3$ and the number of particles to 128^3. Note that the CAMELS simulation are produced in a box of 25 $(h^{-1}\ \mathrm{Mpc})^3$, therefore in this test we take as reference the halofit power spectrum $P(k)/P(k)^{halofit}$. Based on the bottom panel of Figure 4, we can see that, even in this extreme case, the performance of the NN correction remains very good, while PGD no longer applies the desired correction.
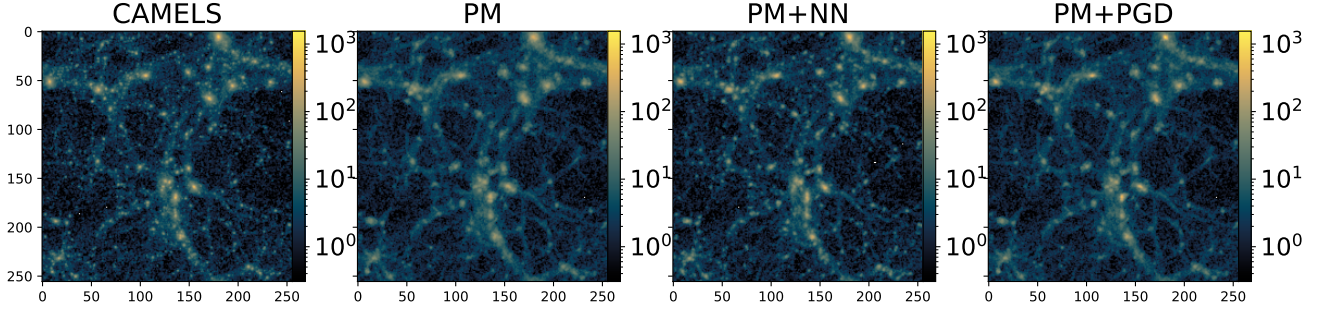
Figure 2. From left to right: Projections of final density field for CAMELS, PM only, PM corrected by the NN scheme and PM corrected by PGD. Although the PGD model improves the sharpness of the halos, it fails to fully recreate the dark matter structures, producing a smoother density profile compared to the NN model, which appears closer to CAMELS.

6. Conclusion

We have presented in this work an alternative to the PGD correction scheme for a quasi N-body PM solver, based on Neural Network implemented as a Fourier-space filter. We illustrated the merits of this approach by comparing the results of a corrected PM simulations against the high resolution CAMELS simulations. We benchmarked our model against the PGD scheme, showing that, at fiducial cosmology, the two methods are able to give similar improvement to the small scale power spectrum, but significantly different improvement in the correlation coefficients. We have also presented the results to changes in resolution and cosmological parameters, proving that our method gives better results for both the power spectrum and cross correlation coefficients, it is therefore less sensitive to the setting of the simulations used for the training. We note however, that our scheme slightly overemphasizes the small scale power for $k \sim 7$ and is not able to strongly improve the results for the cross correlation coefficients when tested to simulations of higher resolutions.
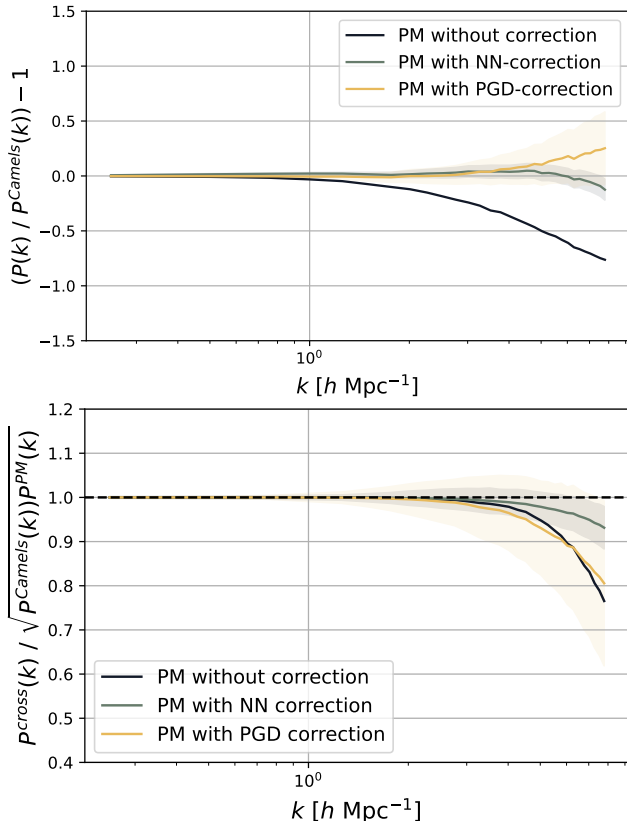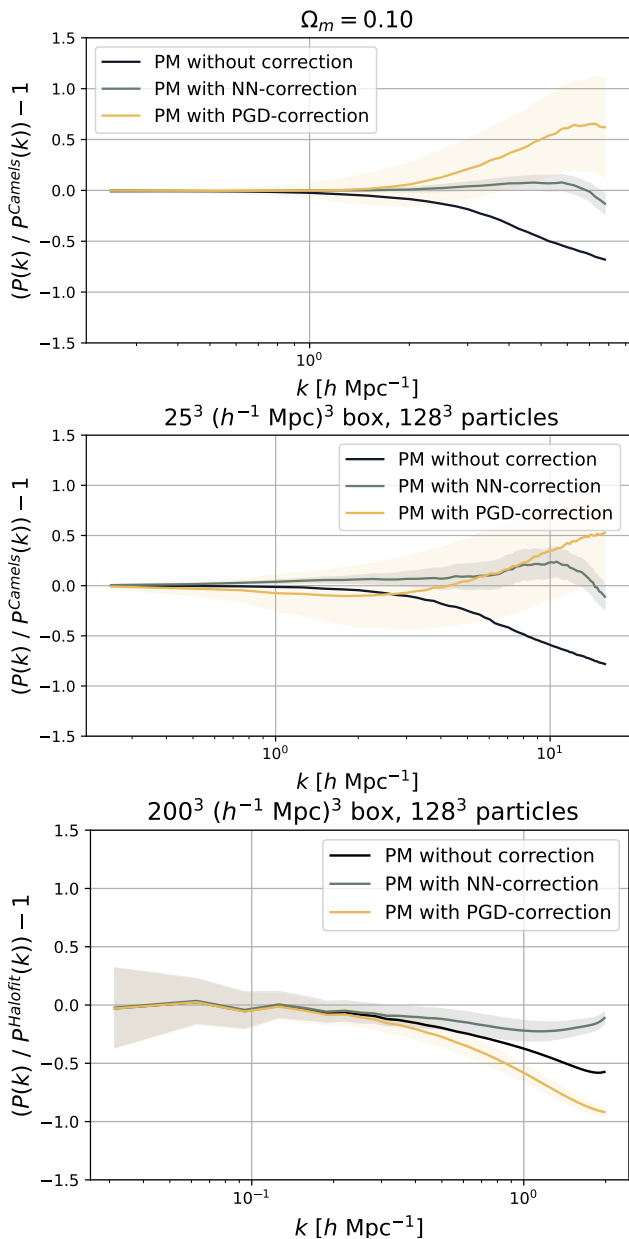

Figure 3. Top panel: Fractional matter power spectrum of PM simulations before and after using the correction models and CAMELS simulations. Bottom panel: cross correlation coefficients with the reference simulation before and after the two different correction schemes. The shaded regions represent the standard deviation from the same realisation corrected with 10 different parameters fitted from different initial conditions.

Figure 4. Benchmarks on matter power spectrum, varying the setting of the simulations. Top panel: effects due to varying the cosmological parameters Ω_m employed in the simulation 🔗. Center panel: effects due to increasing the number of particles (to 128^3) employed in the simulation 🔗. Bottom panel: effects due to increasing the number of particles (128^3) and box size (205 $h^{-3}\mathrm{Mpc}^3$) employed in the simulation. Note that in this last comparison the theoretical HaloFit matter power spectrum is considered as reference 🔗. In the top and central panel the shaded regions represent the standard deviation from the same realisation corrected with 10 different parameters fitted from different initial conditions. In the bottom panel the shaded regions represent the standard deviation from 1000 realisations with different initial conditions corrected with the same parameters.

References

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. JAX: composable transformations of Python+NumPy programs, 2018. URL http://github.com/google/jax.

Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Dai, B., Feng, Y., and Seljak, U. A gradient based method for modeling baryons and matter in halos of fast simulations. *Journal of Cosmology and Astroparticle Physics*, 2018(11):009, 2018.

Feng, Y., Chu, M.-Y., Seljak, U., and McDonald, P. Fastpm: a new scheme for fast simulations of dark matter and haloes. *Monthly Notices of the Royal Astronomical Society*, 463(3):2273–2286, 2016.

He, S., Li, Y., Feng, Y., Ho, S., Ravanbakhsh, S., Chen, W., and Póczos, B. Learning to predict the cosmological structure formation. *Proceedings of the National Academy of Sciences*, 116(28):13825–13832, 2019.

Hockney, R. W. and Eastwood, J. W. *Computer simulation using particles*. crc Press, 2021.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Li, Y., Ni, Y., Croft, R. A., Di Matteo, T., Bird, S., and Feng, Y. Ai-assisted superresolution cosmological simulations. *Proceedings of the National Academy of Sciences*, 118 (19):e2022038118, 2021.

Nelson, D., Springel, V., Pillepich, A., Rodriguez-Gomez, V., Torrey, P., Genel, S., Vogelsberger, M., Pakmor, R., Marinacci, F., Weinberger, R., et al. The illustristng simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):1–29, 2019.

Pontriagin, L. *The Mathematical Theory of Optimal Processes*. International series of monographs in pure and applied mathematics. Pergamon Press; [distributed in the Western Hemisphere by Macmillan, New York], 1964. ISBN 9780080101767. URL https://books.google.fr/books?id=aakrAAAAYAAJ.

Pontryagin, L., Boltyanski, V., Gamkrelidze, R., and Mishchenko, E. The mathematical theory of optimal processes. 1962. interscience. *New York*, 1962.

Tassev, S., Zaldarriaga, M., and Eisenstein, D. J. Solving large scale structure in ten easy steps with cola. *Journal of Cosmology and Astroparticle Physics*, 2013(06):036, 2013.

Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., Spergel, D. N., Somerville, R. S., Dave, R., Pillepich, A., Hernquist, L., Nelson, D., Torrey, P., et al. The camels project: Cosmology and astrophysics with machine-learning simulations. *The Astrophysical Journal*, 915 (1):71, 2021.

A. Train and validation loss for PGD

We fit the PGD parameters only for the final snapshot (a=1) by penalizing the following loss function:

$$\mathcal{L} = \left\| \sigma(k) \left(\frac{p(k)}{p^{Nbody}(k)} - 1 \right) \right\|_2^2, \tag{5}$$

where:

$$\sigma(k) = \frac{1 - \frac{k}{\pi \left(\frac{\text{mesh shape}}{\text{box volume}} \right)}}{\sum_k \left(1 - \frac{k}{\pi \left(\frac{\text{mesh shape}}{\text{box volume}} \right)} \right)} \tag{6}$$

is the weight we add to the loss function to downweight the small scales compared to the large scales. We find that this weight choice gives a better improvement in the small scale power spectrum compared to the only power spectrum case. However it doesn't lead to any improvements or drastic differences to the performance of PGD when applied to simulations with different cosmologies or resolutions.

B. Cross-Correlation

As described in Sec.5, we test the robustness of our correction scheme to changes in resolution and cosmological parameters. In the top panel of Figure 5 we show the results for the cross correlation coefficients between the Camels and the PM field, both computed from simulation generated at the fiducial cosmology of $\Omega_m = 0.10$. It can be seen that, while the cross correlation coefficients between Camels and PM improves after the NN correction, the improvement after the PGD correction is less evident.

The results for the cross correlation coefficients between the Camels and the PM field computed when increased the resolution of the simulation are shown in the bottom panel of Figure 5. We point out that the improvement after our scheme correction is better than PGD, although most of the correlation between the two different fields is not recovered on small scales.
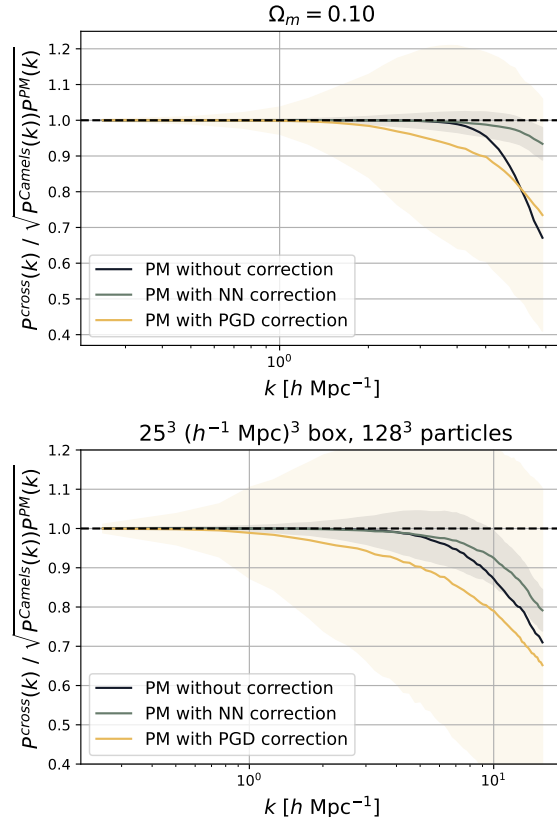

Figure 5. Cross-correlation of PM and CAMELS simulations, before and after using our models and PGD. Top panel: varying the cosmological parameters Ω_m employed in the simulation 🔗. Bottom panel: varying the number of particles (128^3) employed in the simulation 🔗. The shaded regions represent the standard deviation from the same realisation corrected with 10 different parameters fitted from different initial conditions.