SISSA
Scuola
Internazionale
Superiore di
Studi Avanzati

Physics Area - PhD course in
Theoretical and Scientific Data Science

Towards fast and robust Large Scale Structure analysis with physics-based emulators

Candidate:
Mauro Rigo

Advisor:
Roberto Trotta
Co-advisor:
Matteo Viel

Academic Year 2025-26



Abstract

The next generation of large-scale structure surveys will extract unprecedented volumes of cosmological data from the full non-Gaussian structure of the observed fields, requiring fast forward models accurate on non-linear scales and powerful and robust inference pipelines to analyze.

This thesis develops and characterizes a physics-based approach, `JERALD`—JAX Enhanced Resolution Approximate Lagrangian Dynamics—, that produces high-resolution dark matter, stellar mass and neutral hydrogen maps from lower-resolution approximate N-body simulations using only tens of physically-interpretable parameters against the millions to billions of black-box state-of-the-art machine learning approaches. The model displaces the particles from the dark-matter-only cheap simulation via the gradient of an effective potential and, for baryonic outputs, adds a final parametrized field transformation to mimic astrophysical effects that displacement alone cannot capture, producing output maps in excellent agreement with full-hydrodynamical simulations with $8\times$ higher resolution, reproducing the reference power spectrum up to 90% below $k \simeq 1\,h\mathrm{Mpc}^{-1}$ and 70% up to $k \simeq 10\,h\mathrm{Mpc}^{-1}$.

The central part of the thesis asks whether such a model can be extended to depend on cosmological and astrophysical parameters non-trivially, i.e. without optimizing a new set of model parameters for each cosmology, which is what its use inside an inference pipeline would require. The optimized parameters are observed to scatter both for different initial conditions and different optimization initializations even at fixed cosmology, which precludes a direct mapping from cosmology onto them. Nonetheless, they are shown not to be individually determined: rather, the loss constrains only one of them and a combination of the others, a result derived analytically from the structure of the objective and verified across the CAMELS simulation suite. This result reduces the emulation problem from a degenerate parameter vector to one parameter and a single function of scale, and makes it possible to quantify how much of these two objects can be predicted from the properties of a simulation, and why the remainder cannot. The residual dependence is characterized and found to be in part carried by phase information of the input field, but also in part by realization noise comparable in size to the cosmological signal in the volumes available. A single fixed parameter set transfers across cosmologies, reproducing the reference power spectrum to within ten per cent out to $k \simeq 7\,h\mathrm{Mpc}^{-1}$ on simulations it was never fitted to; this, however, is only valid at the level of the mean and two-point correlations, and is therefore not sufficient for field-level inference pipelines. The analysis concludes by identifying the practical and methodological steps that would be required to turn `JERALD` into a usable emulator.

The last part of this thesis applies field-level simulation-based inference to emulated galaxy and neutral hydrogen fields, although using a neural emulator rather than `JERALD`,

finding that preserving the three-dimensional structure of the data and combining the two tracers yields substantially tighter and better-calibrated constraints than two-point compression, with the largest gains obtained when marginalizing over astrophysical uncertainties. Finally, the results of this part indicate how the pipeline would benefit from the physically motivated correction that `JERALD` could introduce.

Contents

Abstract **i**

List of figures **vi**

List of tables **xv**

1 Introduction **1**
1.1 The ΛCDM model . . . 2
1.1.1 The universe over large scales . . . 2
1.1.2 The FLRW metric and Friedmann's equations . . . 2
1.1.3 Hubble's law and the expanding universe . . . 3
1.1.4 Comoving and proper distances . . . 4
1.1.5 Cosmological redshift . . . 4
1.1.6 Matter and energy contents of the universe . . . 6
1.2 Linear perturbation theory and structure growth . . . 9
1.2.1 The Zel'dovich approximation . . . 10
1.2.2 The linear matter power spectrum . . . 11
1.2.3 Baryon acoustic oscillations . . . 12
1.3 Observing the universe . . . 13
1.3.1 Statistical tracers of structure . . . 13
1.3.2 Observational probes . . . 14
1.3.3 The non-linear regime and the need for simulations . . . 16
1.4 Simulations and emulation . . . 16
1.4.1 *N*-body simulations . . . 18
1.4.2 From dark matter to observables . . . 21

2 JERALD: a physics-informed tool for accelerating cosmological simulations **24**
2.1 The model . . . 25
2.2 Neutral hydrogen . . . 29
2.3 Results . . . 32
2.3.1 Code specifications . . . 32

2.3.2 Simulations and training . . . 33
2.3.3 Evaluation of JERALD's output maps . . . 36
2.3.4 Generalization capabilities . . . 40
2.4 Parameters degeneracies . . . 41
2.5 Conclusions . . . 43

3 JERALD as an emulator **45**
3.1 Conditioned JERALD . . . 47
3.1.1 Neural JERALD . . . 47
3.1.2 Spline Green's function . . . 48
3.2 Redshift dependence . . . 52
3.2.1 Redshift evolution of the input and reference simulations . . . 54
3.2.2 Optimization properties and loss landscape . . . 56
3.3 Cosmological and astrophysical inputs . . . 60
3.3.1 Simplified model and training setup . . . 61
3.3.2 Direct regression using a neural network . . . 63
3.3.3 Fixed cosmology: cosmic variance and parameter degeneracies . . . 64
3.3.4 $J(k)$ as training objective . . . 71
3.4 Predicting the model features . . . 72
3.4.1 A physical hypothesis for γ . . . 72
3.4.2 Correlations with the input field . . . 73
3.5 Direct regression of the model features . . . 75
3.5.1 Predicting γ . . . 76
3.5.2 Predicting $J(k)$ . . . 79
3.5.3 Limitations . . . 81
3.6 Performance at fixed parameters . . . 81
3.6.1 Nearest neighbor interpolation . . . 85
3.6.2 The role of γ and $J(k)$ in transferability . . . 86
3.7 Performance limits . . . 87
3.8 Summary and outlook . . . 89

4 Simulation Based Inference on fields **92**
4.1 Introduction . . . 92
4.2 Methodology and results . . . 95
4.2.1 Overview of pipeline . . . 95
4.2.2 Dark matter approximate maps . . . 96
4.2.3 Galaxies and HI maps . . . 98
4.2.4 Simulation-Based Inference with Neural Posterior Estimation . . . 99
4.3 Results . . . 104

4.3.1 Statistical precision . . . 104
4.3.2 Bias and calibration . . . 106
4.3.3 Overall remarks . . . 107
4.4 Emulator . . . 119
4.4.1 Emulator architecture . . . 119
4.4.2 Emulator training . . . 121
4.5 Discussion and conclusions . . . 125

5 Conclusions **127**

A Cloud-In-Cell paint and readout **130**
A.1 CIC calculation of density . . . 130
A.2 CIC readout . . . 130

B Domain decomposition and halo exchange **132**

C Automatic differentiation **134**
C.1 VJP of CIC paint and readout . . . 134
C.2 VJP of Discrete Fourier Transforms (DFTs) . . . 135
C.3 Calculation of the power spectrum and its VJP . . . 136

D Numerical operators: derivatives and MAS **139**
D.1 Spectral and finite-difference derivatives . . . 139
D.2 MAS corrections . . . 140
D.3 Degeneracy with the Green's function . . . 140

E Role of $J(k)$ and γ in the loss **142**

F Calculation of the FT of the CIC kernel **149**

G Calculation of the FT of the sampling function **150**

List of Figures

1.1 Two snapshots of the universe at different times (t_1 in the past and t_0 now). The axes show the physical coordinates, while the grid represent comoving coordinates. The orange segment, highlighting the distance between two comoving points, has constant length d_{com} in comoving coordinates, but its length changes in physical coordinates according to $d_{\text{phys}}(t) = a(t)\, d_{\text{com}}$. Distances are in arbitrary units. 5

1.2 Three snapshots of the universe at different times $t_0 > t_2 > t_1$, showing the stretching of a packet of light emitted from a source at some wavelength λ_{emit} and finally observed at time t_0 (now) with wavelength $\lambda_{\text{obs}} = \lambda_{\text{phys}}(t_3) = \lambda_{\text{emit}}/a(t_1) = 2\lambda_{\text{emit}}$. Correspondingly, the color of the packet of light becomes progressively redder. Source and observer are assumed to have zero peculiar velocities. Distances are in arbitrary units. . . 5

1.3 Evolution over cosmic time of the density parameters Ω_m, Ω_r and Ω_Λ. The plot highlights the crossover points between the three main eras of cosmic history: radiation-dominated at early times, matter-dominated from $z \sim 3400$ to $z \sim 0.3$, and finally dark-energy-dominated until now (and beyond). 7

1.4 A thin slice of the map produced by the DESI five-year survey shows galaxies and quasars above and below the plane of the Milky Way. The Universe's large-scale structure is visible in the magnified inset. Earth lies at the center of the wedges, and the black gap marks where our own galaxy obscures distant objects. Light from the furthest galaxies shown is 11 billion years old by the time it reaches Earth. Credit: Claire Lamman/DESI collaboration 15

1.5 Matter power spectrum at $z = 0$ for the Planck 2018 base-ΛCDM cosmology, with $\sigma_8 \simeq 0.81$. *Top:* linear theory (dashed) compared with the non-linear prediction of the `halofit` prescription (solid). *Bottom:* ratio between the two. 17

2.1 Visual pipeline for how JERALD improves an input DM distribution. The orange box highlights what the L2 loss is calculated over, and the red arrow indicates which parameters are updated via the gradient of the loss. 28
2.2 Same as Figure 2.1 but for stellar mass. The field produced after the displacement layers is an intermediate step generated before the field transformation and cannot be interpreted as any physical field. The stellar mass prediction starts from particle positions already improved by a JERALD pretrained on the DM field. 28
2.3 Same as Figure 2.1 but for neutral hydrogen. The intermediate field generated before the field transformation is again colored in gray for the same reason discussed in Figure 2.2. The figure also highlights the presence of a stellar mass field, pre-computed with JERALD, both in the displacement(s) (Eq. (2.12)) and in the field transformation (Eq. (2.13)). 31
2.4 Time per optimization step (left) and resident set size (RSS, right, a measure of memory used) for JERALD and the original LDL code on one node of the Leonardo supercomputer. The RSS is given by the slurm tool sacct, which may not capture rapid memory usage spikes. Times are given as averages over 20 optimization steps in the case of 512^3 particles and a 512^3 voxels DM target map and 1 displacement layer, and the RSS values are those associated with the 20 steps. The dashed lines in the plot on the left indicate the power law $\propto (N_{\rm CPUs})^{-0.7}$ for reference. The maximum RSS for LDL is not visible as it coincides with the average RSS. 33
2.5 Dark matter overdensity, stellar mass and neutral hydrogen number density maps, from left to right. Each pair of maps compares the reference map from Sherwood-Relics (top panel) with JERALD's output (bottom panel), for $z = 5$, $z = 2$ and $z = 0$ (top to bottom). Notice the different range in the colormaps for neutral hydrogen. All images were obtained as $10h^{-1}$Mpc z-slices of the full $(80h^{-1}\text{Mpc})^3$ simulation box. All voxels used to produce the images were part of either the validation or the test set; the red rectangles in the top left pair of maps—omitted in the others—represent the test subbox, and it extends for the full length of the z-slice used to produce these images. Note that the HI map for $z = 0$ was normalized with the average of the reference map. 35

2.6 Relative difference between $z = 0$ FastPM only and JERALD's output for dark matter, stellar and neutral hydrogen maps. Each map used to compute the relative differences was clipped below 0.5 times the average mass of the reference map and smoothed with a Gaussian filter with a radius of $200h^{-1}$kpc. The plots show the relative difference averaged over the entire length of the $80h^{-1}$Mpc box along the z direction. 37
2.7 As in Fig. 2.6, but showing the relative difference only for the test set, averaged over the $34.4h^{-1}$Mpc of the test sub-box along the z direction. . . 38
2.8 *Left:* DM power spectra (top) and cross correlation coefficients (bottom) for the reference Sherwood-Relics simulation (solid), the approximate N-body FastPM simulation (dotted) and the map improved by JERALD (dashed), for three different redshifts. For $z = 2$ and $z = 5$, the improved power spectra are almost not visible as they overlap with the reference ones. *Right:* Stellar power spectra (top) and cross correlation coefficients (bottom) for the reference Sherwood-Relics simulation (solid) and the map produced using JERALD (dashed) at the same redshifts. 39
2.9 Low-intermediate density neutral hydrogen power spectra (top) and cross correlation coefficients (bottom) for the reference Sherwood-Relics simulation (solid) and the map produced using JERALD (dashed), at three different redshifts. For visualization purposes, the green curves ($z = 5$) in the top panel were offset by multiplying them by an arbitrary constant (1.7), otherwise they would overlap with the orange ones ($z = 2$). The correct scale of the power spectra at $z = 5$ is the same as for $z = 2$. 39
2.10 Similar to Figure 2.5, but for a new Sherwood-Relics simulation unseen by JERALD. In this case, all plots correspond to $z = 0$ and the different quantities—i.e., DM overdensity, stellar mass and HI number density—are arranged horizontally. Here no test sub-box is highlighted, because JERALD was pre-trained on a different simulation. 40
2.11 Dark matter, stellar and neutral hydrogen power spectra at $z = 0$ and $z = 2$ for a target simulation with box size $(80h^{-1}\text{Mpc})^3$ and mesh size 1024^3 with different initial conditions with respect to the one on which JERALD's parameters were trained, along with the results from JERALD. . . 41

2.12 *Left:* training and validation losses for two DM optimizations as discussed in Section 4.3, with a single displacement layer, for two different sets of initial parameters. *Right:* Green's function at convergence for the two set of initial parameters, illustrated in the range of frequencies of the associated maps. The dots on the top-left highlight the very few wavenumbers (10) for which the two curves do not overlap at large scales, whereas there are $\gtrsim 3 \times 10^5$ values of k above $\sim 5h\text{Mpc}^{-1}$ for which the curves significantly differ. 42
2.13 Final power spectra (top), cross correlation coefficients (middle) and power spectra relative difference (bottom) for two DM optimizations as discussed in Section 4.3, but with a single displacement layer, for two different sets of initial parameters. The curves overlap within a few percent. 42

3.1 Pipeline of `JERALD` used as an emulator, together with the linear matter power spectrum generation and the approximate N-body simulator (FastPM). 45
3.2 Training and validation losses (*left*) and cross correlation coefficients (*right*) for the original `JERALD` (OG `JERALD`) with 2 displacement layers and the version with the B-spline with a single displacement layer, both trained in the same setting and on the same Sherwood-Relics target as in Chapter 2. 48
3.3 Power spectra (*left*) and their relative differences with respect to the reference (*right*) for the original `JERALD` (OG `JERALD`) with 2 displacement layers and the version with a single B-spline displacement layer, both trained on the same Sherwood-Relics target. 49
3.4 Bispectra (*left*) and their relative differences (*right*) with respect to the reference for $k_1 = k_2 = 0.5\,h\text{Mpc}^{-1}$ for the original `JERALD` (OG `JERALD`) with 2 displacement layers and the version with the B-spline in the same configuration as for Figures 3.2 and 3.3, both trained on the same Sherwood-Relics target. 50
3.5 Same as Figure 3.4, but for $k_1 = k_2 = 1\,h\text{Mpc}^{-1}$. 51
3.6 Same as Figure 3.4, but for $k_1 = k_2 = 2\,h\text{Mpc}^{-1}$. 51
3.7 *Left:* behavior of the model parameters as functions of redshift, with optimizations starting from the optimized parameters at a previous redshift (e.g. the initial guesses for the parameters at $z = 6$ are the optimized parameters for $z = 7$). *Right:* corresponding Green's functions. 53
3.8 *Left:* cross-correlation coefficients between Sherwood snapshots at $z = 4.2$ and 3.6 and between $z = 3.6$ and 3.2. *Right:* DM power spectra for $z = 8$ down to 3.2 for Sherwood (solid lines) and the approximate N-body simulator (FastPM, dashed lines). 55

3.9 Mean displacement with bars indicating $\pm 1\sigma$, unprocessed (*left*) and corrected to account for the size of the redshift steps (*right*) to show the underlying displacement pattern. Each point represents the average displacement between the indicated redshift and the one to its left (e.g. the point at $z = 10$ represents the mean displacement between snapshots at $z = 10$ and $z = 8$). 55
3.10 Validation losses (with the overall minimum subtracted) through 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4 as functions of training epoch. More optimizations epochs were used for $z = 3.2$ in order to better observe the switching between minima. 57
3.11 Parameter α through 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4 as a function of training epoch. 58
3.12 Parameter k_h through 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4 as a function of training epoch. 58
3.13 Parameter k_l (x axis) vs parameter ν (y axis) throughout 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4. 59
3.14 Green's functions at best validation loss for 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4. 59
3.15 Power spectra relative differences with respect to the Sherwood reference for 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4. The parameters used for the `JERALD` output maps are the ones associated to the best validation losses for every optimization. 60
3.16 Performance of `JERALD` on the CAMELS CV14 simulation with parameters optimized on the CAMELS CV2 simulation with and without validation loss. *Top:* power spectra (left) and cross correlation coefficients (right), *bottom:* bispectra for varying arm lengths $k_1 = k_2$ for an equilateral triangle configuration (left) and for $\theta = \pi/2$ (right). In every case the generalization performance of `JERALD` optimized with validation loss (green lines) is almost the same as that without validation loss (red lines). 63
3.17 Optimized `JERALD` parameters for 11 different CAMELS CV simulations. The x axis simply represents an index of the optimization (35 optimizations with different initial guesses for $\boldsymbol{\theta}$ per CV index). The crosses highlight the parameters associated to the best losses across all 35 optimizations with fixed CV index. 65
3.18 Optimized `JERALD` parameters k_l and ν for 200 optimizations on CV2 starting from different initial random guesses colored by the minimum loss achieved, showing their degeneracy. 66
3.19 Green's functions for the optimized `JERALD` parameters in Figure 3.18. . . . 66

3.20 Optimized Green's functions for all CV optimizations considered. *Left:* colored by CV index, showing the best loss for fixed CV index as a dashed line, *right:* colored by associated γ. 67
3.21 Optimized Green's functions multiplied by $\sqrt{P_{\rho^\gamma}(k)}$, the square root of the power spectrum of the source field ρ^γ (*left*), compared with the Green's functions alone (*right*), for all CV optimizations. 68
3.22 Same as figure 3.21, showing the average curves with the associated spread (1σ, in log space). 68
3.23 Ratio between $\sqrt{P_\rho(k)^\gamma}$ and $\sqrt{P_{\rho^\gamma}(k)}$ across all CV optimizations, colored by γ, before (*left*) and after (*right*) including the correction factor $A_\gamma\, k^{3(\gamma-1)/2}$ of Eq. (3.4). 70
3.24 Scatter plot of γ optimized for 500 LH simulations against $\hat{\sigma}_2$ computed on the FastPM input. The red line was computed by binning both x and y axes into bins of (approximate) equal numbers of points. 74
3.25 $J(k)$ for the optimizations on the LH set introduced in Section 3.3.4 (blue curves) and for the ones on the CV dataset of Section 3.3.3 (orange curves). 80
3.26 Performance of `JERALD` with a single fixed parameter set on 100 simulations it was not optimized on. *Top row*: relative difference between the power spectrum of the output and that of the reference. *Bottom row*: cross-correlation coefficient between output and reference fields. *Left column*: the 100 test simulations of the LH set, comparing the uncorrected FastPM input (orange) with `JERALD` (green). *Right column*: same as above, with the addition of the 500-simulations train set (magenta), in which each simulation carries its own individually optimized parameters. Lines in transparency show individual simulations, thick lines their mean, and shaded areas represent $\pm1\sigma$ regions. 83
3.27 Same as Figure 3.26 for equilateral triangle bispectra (equal arm lengths, $\theta = \pi/3$). 83
3.28 Performance of `JERALD`, in terms of power spectrum relative difference with respect to the reference CAMELS simulation, on 100 simulations it was not optimized on, where each simulations inherits the model parameters of the closest one in the $\boldsymbol{\phi}$ parameter space. *Left*: the distance is computed on the entire $\boldsymbol{\phi}$ vector; *right*: the distance is computed only using the input (Ω_m, σ_8). Similarly to Figures 3.26 and 3.27, lines in transparency show individual simulations, thick lines their mean, and shaded areas represent $\pm1\sigma$ regions. 85

3.29 Generalization performance of the 500 LH train simulations against the recovered value of γ. The vertical axis is the fractional residual power of Eq. (3.11) over a common set of 100 unseen simulations, lower being better. Note that, since every parameter set is scored on the same 100 simulations, differences between points reflect the fitted parameters alone and cannot be attributed to some sets having been evaluated on more difficult boxes. . 86
3.30 Scatter between minimum loss reached in `JERALD` optimizations and the realized input FastPM amplitude $\hat{\sigma}_2$ for 500 LH simulations (see Section 3.3.4) and 11 reference CV simulations (see Section 3.3.3). 88
3.31 *Left:* cross correlation coefficients between FastPM (top) and `JERALD` (bottom) and 500 LH simulations (see Section 3.3.4), colored by the realized FastPM amplitude $\hat{\sigma}_2$. *Right:* scatter between the performance improvement of `JERALD` over its FastPM input quantified via the metric in Eq. (3.11) and the realized FastPM amplitude $\hat{\sigma}_2$ for 500 LH simulations (see Section 3.3.4). 88

4.1 *Top*: sketch of the SBI on power spectra pipeline. Power spectra are extracted from the emulated galaxy and/or HI maps, so that pairs of power spectra and cosmological parameters $\theta_{\mathcal{C}} = \{\Omega_{\rm m}, \sigma_8\}$ constitute the SBI training set (observables fields are reduced to power spectra). Once trained, the SBI flow is able to produce posteriors on $\theta_{\mathcal{C}}$ for validation power spectra. *Bottom*: sketch for SBI on maps pipeline. Pairs of emulated full maps and $\theta_{\mathcal{C}}$ are passed for joint training to a Convolutional Neural Network (CNN) which extracts a latent representation of the data, which is then used to train the SBI normalizing flow. 97
4.2 Visualization of table 4.3 data. 105
4.3 Posteriors on $\{\Omega_{\rm m}, \sigma_8\}$ for $\mathcal{P}^{\rm g\times HI}_{\rm fixedA}$ (red) vs $\mathcal{M}^{\rm g\times HI}_{\rm fixedA}$ (2D, light blue, and 3D, blue) for 200 ΛCDM-Planck realizations test set (ground truth in black). Performing inference on the full 3D maps significantly improves precision. Dots in the lower left plot represent the median values computed from all the posteriors samples. 108
4.4 As in figure 4.3, but for the free astrophysics scenario, with astrophysical parameters marginalized over. 109
4.5 Posteriors on $\{\Omega_{\rm m}, \sigma_8\}$ marginalizing over the free astrophysics case 3D$\mathcal{M}^{\rm g\times HI}$ versus fixed astrophysics 3D$\mathcal{M}^{\rm g\times HI}_{\rm fixedA}$ for 200 validation datasets (ground truth in black). Fixing astrophysics shows smaller uncertainties as expected, but marginalized constraints retain cosmological information while being agnostic to the details of baryonic processes. 110

4.6 Posteriors on $\{\Omega_{\rm m}, \sigma_8\}$ (fixed astrophysics) comparing single-tracers cases $3D\mathcal{M}^{\rm g}_{\rm fixedA}$ (purple) and $3D\mathcal{M}^{\rm HI}_{\rm fixedA}$ (orange) with multi-tracer case $3D\mathcal{M}^{\rm g\times HI}_{\rm fixedA}$ (blue). 111
4.7 As in figure 4.6, but for the free astrophysics scenario. 112
4.8 Distribution of marginal posterior summaries for $\Omega_{\rm m}$ (fixed astrophysics case). Top: violin plots showing the distribution of the posterior mean across 200 realization of the ΛCDM-Planck test set (1,000 samples for each posterior). The distance between the solid line (the median of the all posteriors samples) and the dashed line (the ground truth) can be interpreted as the bias of the inference. Bottom: median value of the marginal posteriors widths of the ΛCDM-Planck test set for 68-95% credible intervals; smaller values indicate higher precision levels. 113
4.9 As in figure 4.8, but for σ_8. 114
4.10 As in figure 4.8, but for the free astrophysics scenario. 115
4.11 As in figure 4.10, but for σ_8. 116
4.12 PP plots for the different data configurations for the fixed astrophysics scenario. Deviations within the gray bands are consistent with random sampling noise from a binomial distribution with 68-95% confidence level. Left panel: $\Omega_{\rm m}$. Right panel: σ_8. 117
4.13 As in figure 4.12 but for the free astrophysics scenario. 117
4.14 Free astrophysics case $3D\mathcal{M}^{\rm g\times HI}$, for an emulated test set of 80 maps with true parameters drawn from their priors for both cosmology and astrophysics. *Top:* each blue point shows the median of 1000 (marginal) posterior samples for a single realization, with ground truth value given by the horizontal axis. Error-bars show the 68% credible interval of the corresponding posterior. *Bottom:* PP plots with parameters drawn from their prior ranges (i.e., Bayesian PP plots, averaging over prior distributions). Deviations within the gray bands are consistent with random sampling noise from a binomial distribution with 68-95% confidence level. 118
4.15 Schematics of the emulation pipeline, from the generation of fast DM maps to the `HGL` emulator. DM maps, generated by means of `FastPM`, are passed to the U-net to generate galaxies or HI, conditioning on astrophysics prescriptions. 119
4.16 Emulator performance on `CAMELS` simulations for $25\,h^{-1}$Mpc box size. The top row compares the `CAMELS` and emulated galaxy number count fields, for an exemplificative map. The bottom left compares the power spectrum residual $\Delta_{\rm Pk}$ between `CAMELS` from the test set and emulated fields, and the bottom right compares the voxel-wise histogram. 123

4.17 As in figure 4.16, but for the emulated HI temperature fields. 123
4.18 As in figure 4.16, but for $(50\,h^{-1}\mathrm{Mpc})^3$ maps, generated with the emulator trained on same resolution $(25\,h^{-1}\mathrm{Mpc})^3$ maps. Test set is the CV `CAMELS` suite. 124
4.19 As in figure 4.17, but for $(50\,h^{-1}\mathrm{Mpc})^3$ maps, generated with the emulator trained on same resolution $(25\,h^{-1}\mathrm{Mpc})^3$ maps. Test set is the CV `CAMELS` suite. 124

A.1 Visual representation of the CIC paint algorithm in 2D. The red particle (with unit mass) contributes to its 4 neighboring mesh vertices based on its distance from them. 131
A.2 Visual representation of the CIC paint algorithm in 2D for a set of particles. 131
B.1 Visualization of the CIC algorithm in the case of a mesh (in light gray) split between two MPI ranks. The light blue plane represents the splitting: mesh points to the left of it belong to rank, say, 1, while mesh points to the right belong to rank 2. The orange particle lies fully within the left slice, thus it does not need to be duplicated: in case it belonged to rank 2, it would only need to be sent to rank 1. The same is true for the violet particle; in this case, the visualization highlights which mesh points the mass of the particle contributes to in case one of its coordinate exceeds the mesh limits and Periodic Boundary Conditions (PBC) apply (in this case, for example, $N_z = 10$ and $L_z = 10$, thus the z coordinates of the mesh points are 0, ..., 9, and the z coordinate of the particle is $x_z = 9.4$). On the contrary, the green particle contributes to mesh points both to the left and to the right of the dividing plane, therefore it needs to be duplicated and distributed to both rank 1 and 2. 133

List of Tables

2.1 `JERALD` parameters for DM, stellar mass and HI targets, at a given redshift. N_p is the number of parameters in each part of the model and $N_p^{\rm tot}$ the total number of parameters for each predicted quantity, with $S_{\rm dm}$, $S_{\rm sm}$ and $S_{\rm HI}$ being the number of displacement layers used for DM, stellar mass and HI respectively (all set to 2 except for $S_{\rm sm} = 3$ at $z = 5$). $N_p^{\rm used}$ is the total the number of parameters used in this work (except for stellar mass at $z = 5$, where $N_p^{\rm used} = 18$ (+10)). For stellar mass and HI, the total number of parameters also includes, in brackets, those used to improve the DM distribution (for both) and to produce the stellar mass map (for HI). . 31

2.2 Initial parameters and their values at convergence for two different optimizations in the case of DM discussed in Section 4.3, but with a single displacement layer. 43

4.1 Summary of the different SBI setups explored in this work grouped by category, and corresponding abbreviations. For example: SBI on power spectrum performed with fixed astrophysics and galaxies as single tracer is referred to as $\mathcal{P}^{\rm g}_{\rm fixedA}$ and SBI on maps performed on 3D cubes of both observables, with free astrophysics is referred to as 3D$\mathcal{M}^{\rm g\times HI}$. 101

4.2 Architecture of the SBI model for 3D field-level inference (model 3D$\mathcal{M}$). The input maps are compressed into a 64-dimensional latent representation via a convolutional encoder, and the posterior over parameters is inferred using NPE with a MAF. Note that $d = 2$ (3) for 2D (3D) field-level inference. 103

4.3 Median Figure of Merit FoM$_m$ across test sets, for the explored SBI scenarios. Values are normalized by 3.23×10^2, corresponding to the FoM$_m$ of $\mathcal{P}^{\rm g\times HI}$ case. 105

4.4 Bias on Ω_m and σ_8 across test sets in [%], for the explored SBI scenarios. We report the percentage difference between ground truth value and median value of the posterior samples across the test set. 107

4.5 Architecture of the HGL emulator. All convolutions (Conv3D) are $3 \times 3 \times 3$ with circular padding. Group Normalization (GN) is used throughout, with FiLM conditioning applied within each residual block, as can be seen in the flow diagram in the last row. For stable learning, the second convolutional layer in each ResNet block is initialized to zero. 119

Chapter 1

Introduction

Cosmology is one of the physical endeavors that address some of the most fundamental questions in nature, seeking to understand the physical laws governing the evolution of the universe, its composition, the distribution of its matter content, and its origin. Over the years, three questions in particular have guided the development of this discipline: how did the universe evolve from the highly uniform state observed in the cosmic microwave background to the complex, multi-scale structure observed today; how is its energy budget distributed and what is the nature of its dark components; and how can all this be modeled accurately and efficiently. In this chapter, the next section describes the standard model of cosmology, the ΛCDM model, which constitutes the main current framework to address the first two questions, introducing the concepts and nomenclature necessary for the rest of the thesis; the third question then motivates the analytical, empirical and numerical methods described in the subsequent sections. The interplay between the three questions is the drive for the analysis and the methods discussed in the reminder of this thesis. Specifically, Chapter 2 is based on [1], of which the author of this thesis is the main author, and presents `JERALD`, an effective displacement model which corrects the output of fast approximate solvers towards a full-hydrodynamical reference and produces stellar mass and neutral hydrogen fields at fixed cosmology. Chapter 3 asks whether such model can be made to depend on cosmological and astrophysical parameters, identifying the quantities the loss of the model actually constrains and how much they can be predicted from the properties of a simulation, highlighting what would be needed to recover the rest. Despite not succeeding in making `JERALD` a proper emulator, Chapter 4 applies field-level simulation-based inference to emulated galaxy and neutral hydrogen fields and quantifies how much is gained by preserving the full three-dimensional structure of the data and by combining tracers. The chapter is based on [2], of which the author of this thesis is third author; the statement of contribution is reported at the beginning of the chapter. Finally, Chapter 5 collects these results and discusses how the two lines of work, `JERALD` and simulation-based inference, might be combined.

1.1 The ΛCDM model

The ΛCDM model describes a universe that is homogeneous and isotropic on large scales, composed of cold dark matter (CDM), baryonic matter, radiation, and a cosmological constant Λ driving the late-time accelerated expansion [3, 4, 5]. All these concepts and components are addressed in detail in the next sections.

1.1.1 The universe over large scales

A foundational assumption of the ΛCDM model is the "cosmological principle", which states that the universe, over large scales, is *homogeneous*, i.e. its properties (density, temperature, curvature) are the same at every point in space at any given cosmic time, and *isotropic*, i.e. it looks the same in every direction from any given point. Observationally, the strongest evidence for the cosmological principle comes from the near-uniformity of the Cosmic Microwave Background (CMB) temperature across the sky, with fluctuations of only about one part in 10^5 [6, 7], and from galaxy surveys, which show the distribution of matter becoming progressively more uniform when averaged over sufficiently large volumes [8, 9]. On smaller scales, the universe is neither homogeneous nor isotropic, as attested by the existence of the very structures (galaxies, clusters, filaments, voids) that are the subject of the following chapters; the cosmological principle is therefore best understood as a statistical statement about the universe on average, valid in a well-defined large-scale limit, rather than a description of its small-scale behavior.

1.1.2 The FLRW metric and Friedmann's equations

The metric describing a universe under such assumptions is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric:

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2d\Omega^2\right] \quad , \tag{1.1}$$

where $a(t)$ is called "scale factor" and k is the spatial curvature, set to zero in the flat ΛCDM model favored by current observations [5]. The FLRW metric is the most general one consistent with a homogeneous and isotropic spatial geometry: the cosmological principle forces all of the metric's time-dependence into the single function $a(t)$, common to all observers, while the spatial part reduces to that of a maximally symmetric 3-space, parametrized entirely by k. The dynamics of $a(t)$ is governed by the Einstein field equations,

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \quad , \tag{1.2}$$

which for the FLRW metric reduce to the Friedmann equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \quad , \tag{1.3}$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \quad , \tag{1.4}$$

where ρ and p are the total energy density and pressure, and G is Newton's constant. In Equations (1.2), (1.3) and (1.4), Λ is the "cosmological constant", a key component in the ΛCDM model, as the name suggests. This object was originally introduced by Einstein in his field equations motivated by the desire to admit a static cosmological solution, balancing the attractive pull of matter with a repulsive term [10]. Its role in the present-day ΛCDM model is different at it will be discussed in full in Section 1.1.6: rather than enforcing a static universe, it is required to source the observed accelerated expansion of the universe.

1.1.3 Hubble's law and the expanding universe

In 1929, Edwin Hubble made a ground-breaking observation: the light received from distant galaxies is systematically shifted towards redder (longer) wavelengths, by an amount growing linearly with their distance [11]; this can be interpreted, for small redshifts, as a recession velocity v via the (non-relativistic) Doppler effect, yielding Hubble's law

$$v = H_0\, d \quad , \tag{1.5}$$

where d is the distance of the observer to the source and the constant of proportionality, H_0, is the Hubble constant, commonly expressed as $H_0 = 100\, h\, \mathrm{km\, s^{-1}\, Mpc^{-1}}$, defining the dimensionless Hubble constant h. Crucially, the fact that galaxies recede with a velocity proportional to their distance, in every direction, indicates that space itself is expanding roughly uniformly, rather than galaxies moving through a static space with some particular velocity distribution centered on earth: Hubble's measurement of $H_0 > 0$ established, for the first time, that the scale factor of Eq. (1.1) is presently increasing, i.e. $\dot{a}(t) > 0$, confirming empirically the non-static solutions to the Einstein field equations already derived on purely theoretical grounds by Friedmann [12, 13] and Lemaître [14], the latter having independently proposed an expanding universe and derived a linear velocity-distance relation two years before Hubble's own measurement. This uniform recession of distant sources, driven by the expansion alone, is referred to as the "Hubble flow"; any additional motion of an object relative to it constitutes its "peculiar velocity".

Yet, Hubble's law as stated in Eq. (1.5) strictly describes the relation between reces-

sion velocity and distance *at the present time*, with a single constant H_0. However, there is no a priori reason for the expansion rate to remain constant throughout cosmic history: whether the universe's expansion accelerates or decelerates over time depends on its total matter-energy content, as will be shown in detail in Section 1.1.6. This motivates generalizing the Hubble constant to a time-dependent quantity, the Hubble *parameter* $H(t)$; this quantity is strictly related to the scale factor $a(t)$, as will be explicitly shown in the next section.

1.1.4 Comoving and proper distances

The expansion of the universe leads to an important distinction between two natural notions of distance between two points, which coincide only at the present epoch: the "comoving" and "proper" distances. For two points *comoving* with the overall expansion (i.e. carried along by it, without additional peculiar velocity), their physical, or proper, separation $d_{\rm phys}$ at time t is related to their comoving separation $d_{\rm com}$ (a fixed coordinate distance, unaffected by the expansion) by

$$d_{\rm phys}(t) = a(t)\, d_{\rm com} \quad , \tag{1.6}$$

directly reflecting the form of the spatial part of the FLRW metric (Eq. (1.1)). The entire time dependence between the two is thus factored into $a(t)$, conventionally normalized to $a_0 = 1$ today, so that proper and comoving distances coincide at the present epoch; comoving distances, customarily expressed in units of h^{-1}Mpc, provide a fixed coordinate grid used throughout this thesis, e.g. to express simulation box sizes and clustering scales. Figure 1.1 shows the relations between proper and comoving distances.

Differentiating Eq. (1.6) at fixed comoving separation gives the recession velocity between two comoving points as $v(t) = \dot{a}(t)\, d_{\rm com}$; defining the Hubble parameter as

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} \quad , \tag{1.7}$$

this relation can then be written as $v(t) = H(t)\, d_{\rm phys}(t)$, recovering Hubble's law (Eq. (1.5)) at the present epoch, with $H_0 \equiv H(t_0)$.

1.1.5 Cosmological redshift

Practically, the scale factor at any epoch is accessed through the relative amount by which the light received from distant sources is shifted towards red, called "redshift": as electromagnetic radiation propagates through an evolving background, its wavelength is stretched by the same factor by which space itself has expanded between emission and

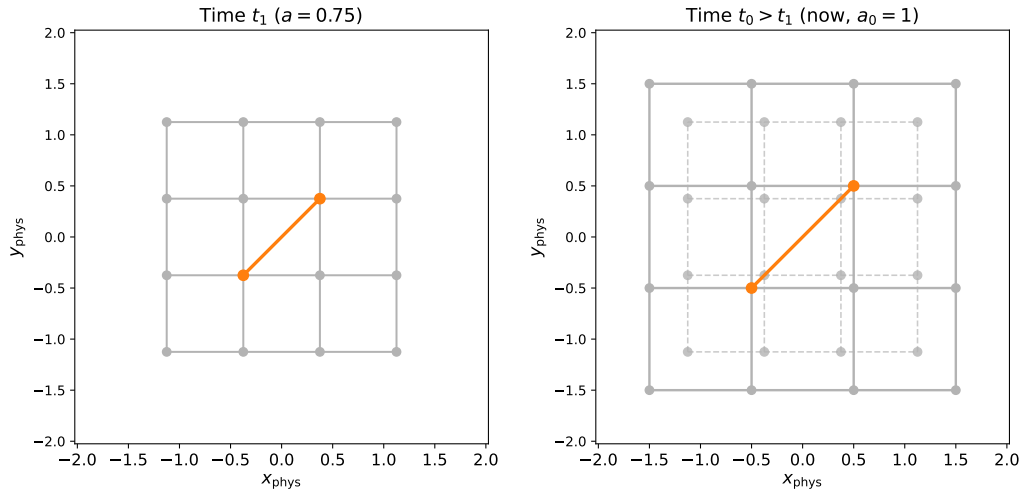


Figure 1.1: Two snapshots of the universe at different times (t_1 in the past and t_0 now). The axes show the physical coordinates, while the grid represent comoving coordinates. The orange segment, highlighting the distance between two comoving points, has constant length d_{com} in comoving coordinates, but its length changes in physical coordinates according to $d_{\text{phys}}(t) = a(t)\, d_{\text{com}}$. Distances are in arbitrary units.

observation. Light emitted with wavelength λ_{emit} at time t_{emit} and observed today with wavelength λ_{obs} is redshifted by

$$z \equiv \frac{\lambda_{\text{obs}} - \lambda_{\text{emit}}}{\lambda_{\text{emit}}} \quad \Rightarrow \quad 1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = \frac{a(t_0)}{a(t_{\text{emit}})} = \frac{1}{a(t_{\text{emit}})} \quad , \tag{1.8}$$

where the last equality uses $a_0 = 1$. Redshift is therefore a direct observable proxy for the scale factor at the time of emission; for this reason, cosmic time is routinely labeled directly by z rather than t, as will be done throughout this thesis, with higher redshift corresponding to earlier cosmic times, and the Big Bang corresponding to $a \to 0$, $z \to \infty$. Figure 1.2 shows the stretching of a packet of light due to the evolving background.

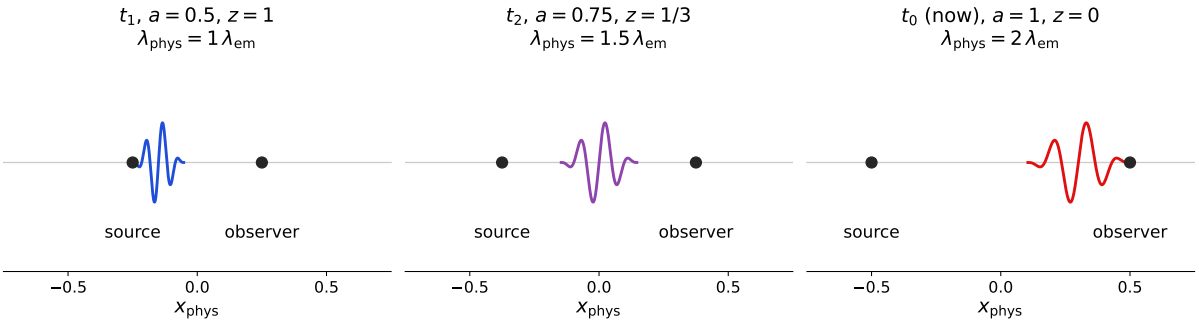


Figure 1.2: Three snapshots of the universe at different times $t_0 > t_2 > t_1$, showing the stretching of a packet of light emitted from a source at some wavelength λ_{emit} and finally observed at time t_0 (now) with wavelength $\lambda_{\text{obs}} = \lambda_{\text{phys}}(t_3) = \lambda_{\text{emit}}/a(t_1) = 2\lambda_{\text{emit}}$. Correspondingly, the color of the packet of light becomes progressively redder. Source and observer are assumed to have zero peculiar velocities. Distances are in arbitrary units.

1.1.6 Matter and energy contents of the universe

Going back to Eq. (1.4), it is convenient to define the critical density

$$\rho_c = 3H^2/8\pi G \tag{1.9}$$

where $H = \dot{a}/a$ is the Hubble parameter. ρ_c is the density required for a spatially flat universe, cf. Eq. (1.3) with $k = 0$: if the total energy density of the universe is higher, $\rho > \rho_c$, then the universe is spatially closed; on the other hand, if $\rho < \rho_c$, then it is spatially open. This definition allows to conveniently express the energy density of each component i (matter, radiation, Λ) in units of ρ_c, defining the density parameters:

$$\Omega_i \equiv \frac{\rho_i}{\rho_c} \quad . \tag{1.10}$$

In particular, $\Omega_m \equiv \Omega_{\rm cdm} + \Omega_b$ is the total matter density parameter (dark matter plus baryons), Ω_r the radiation density parameter (negligible at late times), and Ω_Λ the dark energy density parameter. Flatness implies the constraint

$$\Omega_m + \Omega_r + \Omega_\Lambda = 1 \quad . \tag{1.11}$$

Each of these components redshifts differently with a depending on its equation of state $p = w\rho c^2$: from the conservation of the stress-energy tensor ($\nabla_\nu T^{\mu\nu} = 0$) applied to the FLRW metric, one can derive that, for a perfect fluid, matter ($w = 0$) dilutes as $\rho_m \propto a^{-3}$, radiation ($w = 1/3$) as $\rho_r \propto a^{-4}$, and the cosmological constant ($w = -1$) remains constant, $\rho_\Lambda = \text{const}$. This allows Eq. (1.3) to be rewritten as (remembering also Eq. (1.7))

$$H^2(a) = H_0^2 \left[\Omega_{m,0}\, a^{-3} + \Omega_{r,0}\, a^{-4} + \Omega_{\Lambda,0}\right] \quad , \tag{1.12}$$

which fully determines the expansion history of a flat universe with cosmological constant Λ given the parameters $\{\Omega_{m,0},\, \Omega_{r,0},\, \Omega_{\Lambda,0},\, H_0\}$. The evolution of the density parameters over cosmological time (expressed via redshift) is illustrated in Figure 1.3. In what follows, the subscript 0 indicating the present day density parameters will be dropped for brevity.

Dark matter

Dark matter (DM) is a form of matter that does not emit, absorb, or scatter electromagnetic radiation, and is therefore undetectable through direct observation, interacting with baryonic matter and radiation only gravitationally (or, at most, extremely weakly through other channels). Its existence is only inferred indirectly, from a wide range of independent observations that a universe composed of baryonic matter alone fails to explain. Historically, the first evidence came from the dynamics of galaxy clusters, where Zwicky [15], via

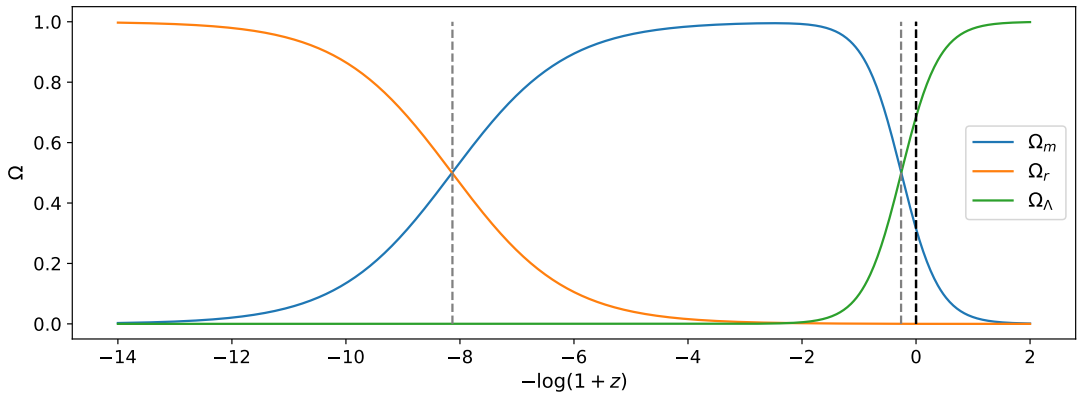


Figure 1.3: Evolution over cosmic time of the density parameters Ω_m, Ω_r and Ω_Λ. The plot highlights the crossover points between the three main eras of cosmic history: radiation-dominated at early times, matter-dominated from $z \sim 3400$ to $z \sim 0.3$, and finally dark-energy-dominated until now (and beyond).

galaxy velocity dispersions, calculated that the Coma Cluster of galaxies possessed about 400 times more mass than could be visually observed; later, systematic measurements of galaxy rotation curves by Rubin and collaborators [16, 17] showed that the rotational velocity of spiral galaxies remains roughly constant at large radii, rather than declining as expected from the visible mass distribution alone, indicating the presence of an unseen mass component enveloping the galaxies. More recently, weak gravitational lensing measurements [18, 19], the amplitude and shape of the CMB power spectrum [5], and Baryonic Acoustic Oscillations (BAOs) [20, 21], discussed in Section 1.2.3, all independently point to the existence of a dominant, non-baryonic matter component. In the ΛCDM model, dark matter is referred to as "cold": this qualifier refers to the kinematic state of DM particles at the time relevant for structure formation, where CDM is assumed to have been non-relativistic. The particle nature of DM remains unknown; candidates commonly discussed in the literature include weakly interacting massive particles (WIMPs), axions, and sterile neutrinos [22].

Baryonic matter and radiation

Baryonic matter comprises the ordinary matter made of protons, neutrons, and electrons[1], i.e. the constituents of stars, gas, dust, and, ultimately, of all directly observable astrophysical objects. Its cosmic abundance is tightly constrained both by Big Bang nucleosynthesis [23] and by the relative heights of the acoustic peaks in the CMB power spectrum [5], and is found to be significantly subdominant with respect to the total matter density, im-

[1] This nomenclature is standard in cosmology. Technically, electrons are not baryons but rather leptons; nonetheless, since the electron mass is smaller than the proton mass by a factor of nearly two thousand, their contribution to the mass density is negligible and the terminology is accurate for all practical purposes.

plying that the majority of matter in the universe is indeed non-baryonic (dark) matter. Radiation refers to the relativistic species in the universe, primarily photons (the CMB itself) and, to a good approximation on cosmological scales, neutrinos before becoming non-relativistic. As mentioned above, radiation redshifts faster than matter, $\rho_r \propto a^{-4}$ as opposed to $\rho_m \propto a^{-3}$, as a consequence of both the dilution of particle number density and the redshifting of individual photon energies; this implies that, despite being negligible today, $\Omega_r \sim 10^{-4}$ [5, 24], radiation dominated the energy budget of the universe at sufficiently early times (radiation domination), with important consequences for the processing of perturbations encoded in the transfer function $T(k)$ of Eq. (1.28).

The cosmological constant and cosmic acceleration

As mentioned in Section 1.1.2, the cosmological constant Λ was originally introduced by Einstein as an additional term in his field equations specifically to allow for a static universe, by balancing the attractive pull of matter with a repulsive term [10]. Following Hubble's observation that galaxies recede with velocities proportional to their distance [11], indicating that the universe is in fact expanding rather than static, the cosmological constant was largely abandoned, as it was no longer needed to enforce a static solution. Interest in Λ was revived at the end of the twentieth century, when observations of distant Type Ia supernovae [25, 26] showed that the expansion of the universe is not just ongoing, but rather accelerating. From the second Friedmann equation (Eq. (1.4)), this requires a component with sufficiently negative pressure, $p < -\rho c^2/3$, to overcome the decelerating effect of matter and radiation, whose positive pressure (or pressureless nature) always contributes with a negative sign to $\ddot{a}$. A cosmological constant is equivalent, in this context, to a fluid with constant energy density and equation of state $p_\Lambda = -\rho_\Lambda c^2$, as mentioned in Section 1.1.6, for which the acceleration equation gives

$$\left.\frac{\ddot{a}}{a}\right|_\Lambda = \frac{\Lambda c^2}{3} > 0 \quad . \tag{1.13}$$

Given the faster dilution with a of this component, it is guaranteed to dominate the energy budget of the universe over matter and radiation at sufficiently late times, thus providing a source for its accelerated expansion. The physical origin of Λ, whether a true constant (e.g. vacuum energy) or a dynamical field with $w \approx -1$ (dark energy), remains an open question and an active area of research [27, 28], but is not the focus of this thesis, which treats Λ as a constant throughout, consistent with the flat ΛCDM model; for references on Λ, see [29, 30, 31].

1.2 Linear perturbation theory and structure growth

The FLRW metric of Eq. (1.1) describes a perfectly homogeneous universe, whereas the observed universe exhibits structure across a wide range of scales, which grew from small primordial density fluctuations and amplified over cosmic time by gravitational instability. As long as these fluctuations remain small, their evolution can be described by linear perturbation theory. The matter density field is described in terms of a small perturbation in the density field, the dimensionless overdensity

$$\delta(\boldsymbol{x},\, t) \equiv \frac{\rho(\boldsymbol{x},\, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \quad , \tag{1.14}$$

where $\bar{\rho}(t)$ is the mean matter density at time t and $\boldsymbol{x} = \boldsymbol{r}/a$ are comoving coordinates, $\boldsymbol{r}$ being physical coordinates. In the linear regime, $|\delta| \ll 1$, and the fluid equations (continuity, Euler, and Poisson) for an ideal pressureless fluid in an expanding background can be combined into a single second-order differential equation for δ (see, e.g., [3]):

$$\ddot{\delta}(\boldsymbol{x},\, t) + 2H(t)\dot{\delta}(\boldsymbol{x},\, t) = 4\pi G\bar{\rho}(t)\, \delta(\boldsymbol{x},\, t) \quad . \tag{1.15}$$

The general solution of such an equation admits a growing and a decaying mode:

$$\delta(\boldsymbol{x}, t) = A(\boldsymbol{x})D_1(t) + B(\boldsymbol{x})D_2(t) \quad ; \tag{1.16}$$

the growing-mode solution can be written in terms of the linear growth factor $D(a)$, such that

$$\delta(\boldsymbol{x},\, a) = D(a)\, \delta(\boldsymbol{x},\, a_i) \quad , \tag{1.17}$$

where $\delta(\boldsymbol{x},\, a_i)$ is the density perturbation at some initial time t_i and $D(a)$ is normalized such that $D(a_i) = 1$. The growth factor depends on the background expansion history, and therefore on Ω_m and Ω_Λ, through Eq. (1.12), and it can be expressed as the integral form of the solution to Eq. (1.15):

$$D(a) \propto H(a) \int_0^a \frac{da'}{\left[a' H(a')\right]^3} \quad . \tag{1.18}$$

It is often useful to define the logarithmic growth rate $f(a) \equiv d\ln D / d\ln a$, which enters, for instance, the mapping between real- and redshift-space clustering through redshift-space distortions [32, 33] (see Section 1.3.1).

1.2.1 The Zel'dovich approximation

Substituting the ansatz in Eq. (1.17) into the Poisson equation for the peculiar gravitational potential Φ[2],

$$\nabla^2\Phi(\boldsymbol{x},\, a) = 4\pi G\bar{\rho}(a)\, a^2\, \delta(\boldsymbol{x},\, a) \quad , \tag{1.19}$$

and using $\bar{\rho}(a) = \bar{\rho}_0\, a^{-3}$ (matter domination, see Section 1.1.6), yields

$$\nabla^2\Phi(\boldsymbol{x}, a) = 4\pi G\bar{\rho}_0\, [D(a)/a]\, \delta_i(\boldsymbol{x}) \quad , \tag{1.20}$$

solved by

$$\Phi(\boldsymbol{x},\, a) = \frac{D(a)}{a}\, \Phi_i(\boldsymbol{x}) \quad , \qquad \nabla^2\Phi_i(\boldsymbol{x}) \equiv 4\pi G\bar{\rho}_0\, \delta_i(\boldsymbol{x}) \quad . \tag{1.21}$$

Notably, in an Einstein-de Sitter universe, where $D(a) \propto a$ [3], Eq. (1.21) shows that the peculiar potential is constant in time: gravitational potential wells neither grow nor decay during matter domination, only the density contrast does. The peculiar velocity, $\boldsymbol{v} = a\dot{\boldsymbol{x}}$, obeys the linearized, pressureless Euler equation

$$\frac{d\boldsymbol{v}}{dt} + H\boldsymbol{v} = -\frac{1}{a}\nabla\Phi \quad , \tag{1.22}$$

which can be rewritten as $d(a\boldsymbol{v})/dt = -\nabla\Phi(\boldsymbol{x},\, a)$. Using Eqs. (1.21) and (1.15), one can find that this integrates exactly to

$$\boldsymbol{v}(\boldsymbol{x},\, a) = -\frac{a\,\dot{D}(a)}{4\pi G\bar{\rho}_0}\, \nabla\Phi_i(\boldsymbol{x}) \quad , \tag{1.23}$$

with the integration constant set to zero, selecting the growing-mode solution.

Since $\boldsymbol{v} = a\dot{\boldsymbol{x}}$, Eq. (1.23) gives $\dot{\boldsymbol{x}} = -[\dot{D}(a)/4\pi G\bar{\rho}_0]\, \nabla\Phi_i(\boldsymbol{x})$. To leading order, the Eulerian position on the right-hand side can be replaced by the fixed, unperturbed Lagrangian position $\boldsymbol{q}$ of the mass element under consideration, so the equation integrates directly in time to

$$\boldsymbol{x}(\boldsymbol{q},\, a) = \boldsymbol{q} + D(a)\, \boldsymbol{\Psi}(\mathbf{q}) \quad , \qquad \boldsymbol{\Psi}(\boldsymbol{q}) \equiv -\frac{\nabla_{\boldsymbol{q}}\Phi_i(\boldsymbol{q})}{4\pi G\bar{\rho}_0} \quad , \tag{1.24}$$

having set the integration constant to zero, consistent with mass elements starting from their unperturbed positions at early times ($D \to 0$). By construction, $\nabla_{\boldsymbol{q}} \cdot \boldsymbol{\Psi}(\boldsymbol{q}) = -\delta_i(\boldsymbol{q})$: the displacement field is sourced directly by (minus) the initial density field.

Eq. (1.24) is the Zel'dovich approximation [34]: a fully Lagrangian description of struc-

[2] The gravitational potential is generally split as $\Phi_{tot} = \Phi_0 + \Phi$, where Φ_0 is the potential of the smooth background and Φ the *peculiar* potential, sourced entirely by the density perturbation δ. Since Φ_0 has no gradient for a perfectly homogeneous background, it drops out of the dynamics entirely, and only Φ is relevant in this section.

ture formation in the linear regime, specifying the trajectory of each mass element solely in terms of its initial position $\boldsymbol{q}$ and a single, time-independent displacement field $\boldsymbol{\Psi}(\boldsymbol{q})$, with all time-dependence factored into the growth factor $D(a)$. Combining Eqs. (1.23) and (1.24) gives the peculiar velocity directly in terms of the displacement itself,

$$\boldsymbol{v}(\boldsymbol{q},\, a) = a\, H(a)\, f(a)\, \big[\boldsymbol{x}(\boldsymbol{q},\, a) - \boldsymbol{q}\big] \quad , \tag{1.25}$$

using $\dot{D} = HfD$: the peculiar velocity of a mass element is directly proportional to its own displacement, with proportionality constant aHf.

Although derived here for a pressureless fluid in the strictly linear regime ($\delta \ll 1$), the Zel'dovich approximation remains, empirically, a remarkably good description of large-scale structure well into the mildly non-linear regime, and constitutes both the leading-order term of Lagrangian perturbation theory and the standard method for generating the initial conditions of N-body simulations, both further discussed in Section 1.4.1.

1.2.2 The linear matter power spectrum

Since the initial density perturbations are stochastic, they are most naturally described statistically rather than deterministically. Under the standard assumption that the primordial fluctuations are Gaussian random fields (as predicted by simple single-field inflation [35, 36, 37]), all statistical information is encoded in the two-point correlation function:

$$\xi(r) \equiv \langle \delta(\boldsymbol{x})\, \delta(\boldsymbol{x} + \boldsymbol{r}) \rangle \quad , \tag{1.26}$$

or equivalently, in Fourier space, the power spectrum $P(k)$, defined via

$$\langle \tilde{\delta}(\boldsymbol{k})\, \tilde{\delta}^*(\boldsymbol{k}') \rangle = (2\pi)^3\, \delta_D(\boldsymbol{k} - \boldsymbol{k}')\, P(k) \quad , \tag{1.27}$$

where δ_D is the Dirac delta function, $\tilde{\delta}(\boldsymbol{k})$ is the Fourier transform of $\delta(\boldsymbol{x})$ and statistical isotropy has been assumed, so that P depends only on the modulus $k = \|\boldsymbol{k}\|$. The linear matter power spectrum at some scale factor a can be written as

$$P(k,\, a) = D^2(a)\, T^2(k)\, P_{\mathrm{prim}}(k) \quad , \tag{1.28}$$

where $P_{\mathrm{prim}}(k) \propto k^{n_s}$ is the primordial power spectrum generated during inflation, with spectral index n_s [38], and $T(k)$ is the transfer function, encoding the scale-dependent processing of perturbations that enter the horizon before matter-radiation equality and are subject to physical effects such as baryonic acoustic oscillations [39, 40], discussed in the next section. The transfer function and the growth factor are typically computed with a Boltzmann solver such as CLASS [41] or CAMB [42], given a set of cosmologi-

cal parameters $\{\Omega_m,\, \Omega_b,\, h,\, n_s,\, ...\}$; this is the "linear power spectrum" input referenced throughout this thesis whenever an approximate N-body solver requires initial conditions. A commonly used normalization for the amplitude of the power spectrum is σ_8, defined as the root-mean-square of the matter density field smoothed with a top-hat filter of radius $8h^{-1}$Mpc,

$$\sigma_8^2 \equiv \frac{1}{2\pi^2} \int dk\, k^2\, P(k)\, W^2(kR)\Big|_{R=8\,h^{-1}\mathrm{Mpc}} \quad , \tag{1.29}$$

where $W(kR)$ is the Fourier transform of the top-hat window function. The scale $8h^{-1}$Mpc is historically chosen as it is roughly the scale at which the variance of the (linearly extrapolated) density field is of order unity, making σ_8 a convenient, scale-independent summary of the clustering amplitude, alongside Ω_m, one of the primary parameters used to characterize a given cosmology in this thesis.

1.2.3 Baryon acoustic oscillations

As mentioned above, Baryonic Acoustic Oscillations (BAOs) represent an important feature of the transfer function. Before recombination, the universe was a plasma of electrons, protons and neutrons, and photons were unable to travel freely, being tightly coupled with the plasma by Thomson scattering. The resulting fluid's pressure was dominated by radiation, with sound speed $c_s \simeq c/\sqrt{3}$. Overdensities in this fluid could not simply grow under gravity: rather, the competing effect of gravitational attraction and pressure generated spherical acoustic waves outwards carrying baryons and photons with it [43, 44], while dark matter, being pressureless and interacting only gravitationally, remained behind. As photons decoupled at the drag epoch, at $z_\mathrm{d} \simeq 1060$ [40], the pressure support disappeared and the wave stalled, leaving the baryons in spherical shells of comoving radius equal to the sound horizon at that time,

$$r_\mathrm{d} = \int_{z_\mathrm{d}}^{\infty} \frac{c_s(z)}{H(z)}\, dz \;\simeq 147\,\mathrm{Mpc} \simeq 100h^{-1}\mathrm{Mpc} \quad . \tag{1.30}$$

where $c_s(z)$ is the sound speed at redshift z. The imprint of this process on the late-time matter distribution is an excess of clustering at separations $r \simeq r_\mathrm{d}$, appearing as a localized peak in the correlation function of Eq. (1.26) and, equivalently, as a series of damped oscillations in $P(k)$ (Eq. (1.27), see also Figure 1.5 for an example) with spacing $\Delta k \simeq 2\pi/r_\mathrm{d} \simeq 0.06\,h\mathrm{Mpc}^{-1}$ [45]. Their amplitude is modest, of order a few per cent, because only the baryons, constituting roughly a sixth of the total matter density, participate, while the dark matter contributes a smooth component that dilutes the feature. The oscillations are nonetheless clearly visible in $T(k)$, and are the reason accurate transfer functions require either a Boltzmann solver or a dedicated fitting formula [40], a smooth power law being unable to reproduce them. Because r_d is fixed by pre-

recombination physics alone, this feature also provides the most robust geometrical probe available to large-scale structure surveys, as discussed in Section 1.3.1.

1.3 Observing the universe

Cosmology occupies a unique position among the physical sciences: the existence of a single realization of the universe, together with the immense scales involved, precludes controlled laboratory experiments on its origin, its evolution and many of its most extreme phenomena. The accuracy of a cosmological theory can therefore be assessed only through its ability to reproduce observations, and the progress of the field is driven by a correspondingly tight interplay between data collection and the theoretical modeling of the observed quantities. On one side, advances in modeling tools and statistical methods extend the range of questions that a given survey can address, and inform the design and optimization of future ones; on the other, increasingly ambitious surveys have produced, and will continue to produce, unprecedented volumes of high-quality data, whose interpretation demands predictions that are both accurate enough to match them and, crucially, efficient enough to be generated thousands of times over. It is also important to stress that what these surveys map is the matter distribution at late times, long after the linear description of Section 1.2 has ceased to hold: the observed field is the evolved, non-linear one, and this is precisely what makes its theoretical modeling difficult, as discussed in Section 1.3.3.

1.3.1 Statistical tracers of structure

This evolved, non-linear field is commonly referred to as the *cosmic web*: a network of filaments and sheets connecting dense clusters and surrounding large, roughly empty voids. It was first mapped observationally in galaxy redshift surveys such as the CfA Redshift Survey [46, 47] and has since been characterized with steadily increasing precision and volume.

While the power spectrum of Eq. (1.27) is defined for the underlying matter density field δ, matter itself is not directly observable; large-scale structure (LSS) is instead accessed through tracers, either discrete, such as galaxies, or continuous, such as the neutral hydrogen (HI) they contain or the gravitational lensing of background light. Such tracers are generally *biased*, in the sense that their density contrast does not coincide with that of the matter field. At leading order this is captured by a linear bias parameter b,

$$\delta_{\rm tracer}(\boldsymbol{x}) = b\,\delta(\boldsymbol{x}) \quad , \tag{1.31}$$

so that $P_{\rm tracer}(k) = b^2 P(k)$ on large scales, where the linear approximation holds. On

smaller, non-linear scales bias becomes scale-dependent and stochastic [48, 49], complicating the translation between observed tracer clustering and the underlying matter distribution. A further complication arises because galaxy positions along the line of sight are inferred from redshifts, which receive a contribution from peculiar velocities in addition to the Hubble flow. The resulting anisotropy of the observed clustering is known as redshift-space distortion (RSD), and on large scales it is well described by the Kaiser formula [32],

$$P_s(k,\mu) = \left(b + f\mu^2\right)^2 P(k) \quad , \tag{1.32}$$

where P_s is the redshift-space power spectrum, μ is the cosine of the angle between $\boldsymbol{k}$ and the line of sight, and f is the logarithmic growth rate introduced in Section 1.2. Equation (1.32) shows that RSD measurements provide direct access to f, and therefore to the growth rate of structure, independently of the amplitude constraint provided by σ_8. In configuration space, the same two-point information is equivalently encoded in the correlation function of Eq. (1.26): historically, $\xi(r)$ was the first statistic used to quantify galaxy clustering [50, 51, 52], and it remains widely used alongside $P(k)$, particularly for isolating localized features such as the BAO peak of Section 1.2.3, which appears in $\xi(r)$ as a single excess at $r \simeq r_{\rm d}$ rather than as the series of damped oscillations it produces in $P(k)$. Since $r_{\rm d}$ is set entirely by pre-recombination physics, it is calibrated to sub-per cent accuracy by CMB observations independently of the late-time expansion history, thus acting as a standard ruler for cosmology [53, 54]. Since the feature sits at large, quasi-linear scales, it is relatively insensitive to galaxy bias and to the details of non-linear evolution.

1.3.2 Observational probes

Different surveys access LSS through different tracers, each with its own advantages and systematics; below are listed and briefly described the most common.

Galaxy redshift surveys

Spectroscopic surveys such as SDSS/BOSS/eBOSS [55, 56, 57] and, more recently, DESI [58] measure galaxy positions and redshifts directly, enabling three-dimensional clustering measurements ($P(k)$, $\xi(r)$, RSD, BAO). Their main limitations are galaxy bias, Eq. (1.31), and the comparatively high cost of obtaining spectroscopic redshifts for large samples.

Weak gravitational lensing

Imaging surveys such as KiDS [59], DES [60] and HSC [61], Euclid [62] and LSST [63], measure the coherent distortion of background galaxy shapes by foreground matter along

the line of sight [64, 65] for lensing. Since lensing is sourced by the total matter distribution, it is free of the galaxy bias uncertainty affecting clustering measurements, at the cost of lower signal-to-noise and of sensitivity to the redshift distribution of the source galaxies.

21-cm intensity mapping

Radio surveys, for instance with SKA [66] or its pathfinders, measure the integrated 21-cm emission from neutral hydrogen [67, 68, 69] without resolving individual galaxies, providing an efficient tracer of LSS particularly at high redshift.

The Lyα forest

Absorption features imprinted by neutral hydrogen along quasar lines of sight trace the matter distribution in the intergalactic medium, probing LSS at high redshift and on comparatively small scales [70, 71, 72].

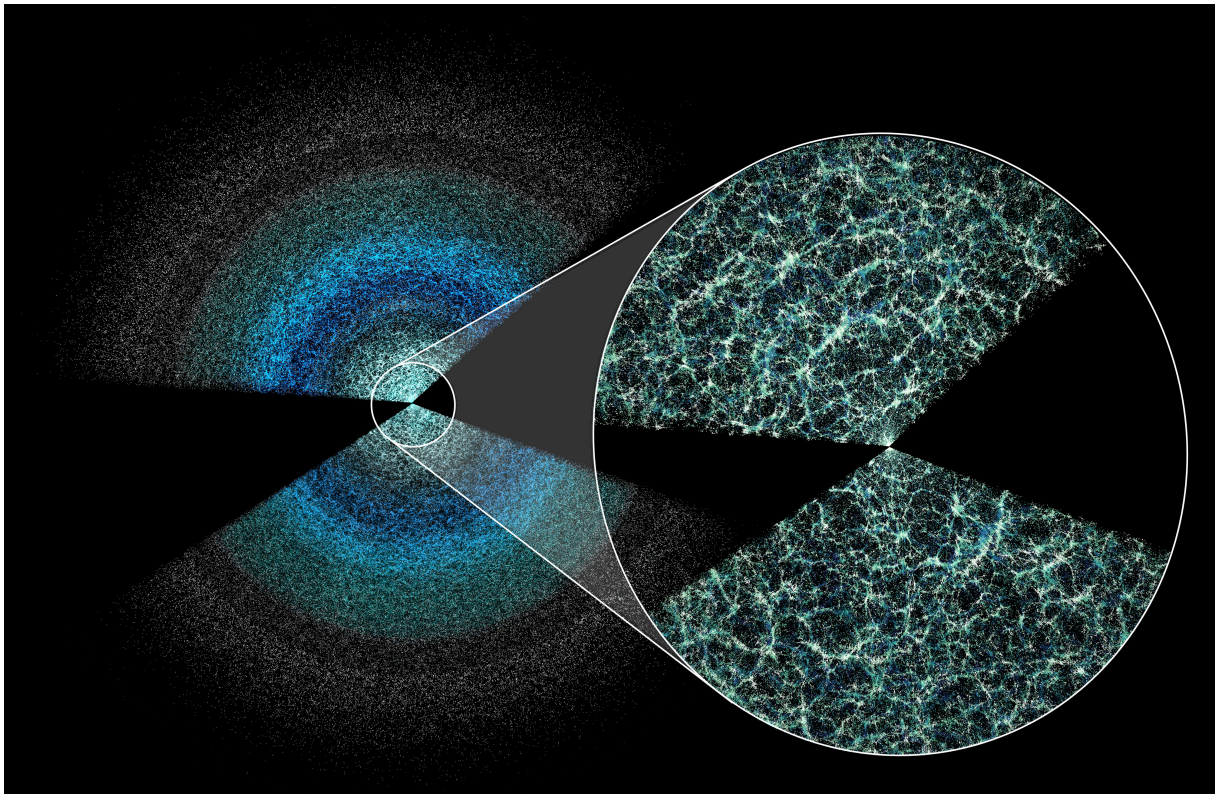


Figure 1.4: A thin slice of the map produced by the DESI five-year survey shows galaxies and quasars above and below the plane of the Milky Way. The Universe's large-scale structure is visible in the magnified inset. Earth lies at the center of the wedges, and the black gap marks where our own galaxy obscures distant objects. Light from the furthest galaxies shown is 11 billion years old by the time it reaches Earth. Credit: Claire Lamman/DESI collaboration

1.3.3 The non-linear regime and the need for simulations

As hinted at above, the linear theory of Section 1.2 describes the density field accurately on large scales and at early times, where $|\delta| \ll 1$. However, at low redshift and on the scales relevant to galaxy formation and to most observational probes of LSS, i.e. typically below $\sim 10h^{-1}$Mpc, density perturbations grow to order unity and beyond, and the linear approximation of Eq. (1.15) breaks down. The evolved field on these scales is strongly non-Gaussian, so that its statistical description requires higher-order correlation functions in addition to the power spectrum, the latter alone no longer capturing all of the available cosmological information. In this regime no general analytic solution to the equations of gravitational instability exists. In this context, one useful statistic, one order higher than the power spectrum, is the bispectrum:

$$\begin{aligned} \langle \tilde{\delta}(\boldsymbol{k}_1)\, \tilde{\delta}(\boldsymbol{k}_2)\, \tilde{\delta}(\boldsymbol{k}_3) \rangle &= (2\pi)^3\, \delta_D(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3)\, B(\boldsymbol{k}_1,\, \boldsymbol{k}_2) \\ &= (2\pi)^3\, \delta_D(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3)\, B(k_1,\, k_2,\, \theta) \quad , \end{aligned} \tag{1.33}$$

where θ is the angle between $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$. This quantity will often appear throughout this thesis. An illustration of the difference between the linear and non-linear power spectrum is provided in Figure 1.5, where the two curves were obtained using CAMB [42]. Note that, for dark-matter-only ΛCDM, `halofit`, a fitting formula motivated by the halo model and calibrated on dark-matter-only N-body simulations [73] and here in the recalibrated form of [74], is accurate at the level of a few per cent up to $k \sim 1\, h\mathrm{Mpc}^{-1}$, degrading to of order ten per cent towards $k \sim 10\, h\mathrm{Mpc}^{-1}$. BAOs (see Section 1.2.3) are visible as a faint modulation to the non-linear spectrum between $k \simeq 0.05$ and $0.35\, h\mathrm{Mpc}^{-1}$, highlighted by a vertical band in the plot. The two spectra agree by construction on large scales and diverge above $k \simeq 0.1\, h\mathrm{Mpc}^{-1}$, the non-linear spectrum exceeding the linear one by more than an order of magnitude by $k \sim 10\, h\mathrm{Mpc}^{-1}$. It is this regime, inaccessible to perturbative treatments and containing most of the information targeted by forthcoming surveys, that requires numerical simulation.

1.4 Simulations and emulation

In this context, numerical simulations have become an indispensable tool, allowing researchers to investigate the nonlinear evolution of cosmic structures, explore the implications of different physical assumptions, and establish direct links between theoretical predictions and observational data. The most advanced simulations, called full-hydrodynamical simulations owing to their treatment of baryons as fluids, are capable of tracking the evolution of several billions of dark matter (DM) and baryonic particles for several billions of years and implement many of the physical phenomena that govern

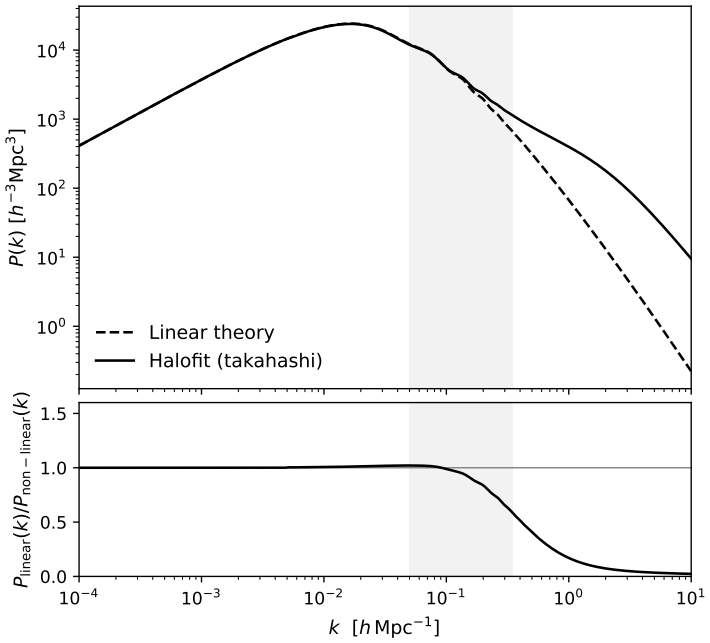


Figure 1.5: Matter power spectrum at $z = 0$ for the Planck 2018 base-ΛCDM cosmology, with $\sigma_8 \simeq 0.81$. *Top:* linear theory (dashed) compared with the non-linear prediction of the `halofit` prescription (solid). *Bottom:* ratio between the two.

their dynamics on a wide range of scales. These simulations can model many different observables, allowing to test descriptions of the universe under different cosmologies, e.g. different and/or exotic DM models or different cosmological parameters, which is essential for deriving constraints or ruling out hypotheses. There exist several codes and a vast zoo of simulation suites with different properties and purposes; among them, for example, the IllustrisTNG suite [75], based on the AREPO code [76], aimed towards the study of galaxy formation and evolution and implementing highly sophisticated treatments for gas cooling, star formation and feedback, features some of the largest simulations to date, while the Sherwood [77] and Sherwood-Relics [78] simulations, based on GADGET-3 (an evolution of GADGET-2 [79]) and focused towards modeling the intergalactic medium during and after the cosmic reionization of hydrogen, implement a vast range of cosmological parameters and a variety of DM scenarios.

However, the level of complexity and accuracy that these simulations manage to achieve comes with an extreme computation cost, often requiring millions of CPU hours to run on significant fractions of entire supercomputers, or up to several tens of days in real time [77, 80]. At the same time, current and future sky surveys continue to produce increasingly large volumes of data: as an example, Euclid will cover more than a third of the sky in its 6-years mission, with angular resolutions up to few tenths of arcseconds, while LSST, in 10 years, will study over 40% of the sky with similar resolution. Therefore, the exciting opportunities that these surveys offer then pose a major problem: the most

sophisticated tools to exploit the full statistical content of the fields observed by cosmological surveys, which can even combine several probes so that their distinct systematics can be cross-checked to break parameter degeneracies, such as Simulation Based Inference (SBI, introduced and discussed in Chapter 4), can require thousands of simulations, rendering the use of these full-hydrodynamical codes practically impossible for this specific purpose. For this reason, several approaches have been proposed to overcome this major bottleneck.

1.4.1 N-body simulations

One of the major culprits for the enormous computational cost of numerical simulations is baryonic matter, since baryons are characterized by a vast range of physical processes that need to be modeled in order to accurately describe their interactions, such as cooling, star formation, feedback, turbulence, etc. [81, 82, 83]. On the contrary, DM is relatively simple to simulate, as it only interacts through gravity; therefore, since gravity is the dominant component over large scales, where baryons trace the evolution of DM, one possible approach for mitigating the great computational cost of full-hydrodynamical simulations consists in running DM-only (N-body) simulations, and then populating them with baryons or extracting baryonic properties through some analytical, empirical or other approximate method in a post-processing step.

Nonetheless, computing forces between billions of particles is still computationally highly expensive, and many different methods have been proposed to amortize this cost as well. The most common is the so-called tree approach [84], where the contribution to the gravitational potential from distant particles is approximated by the lowest order terms in a multipole expansion of their mass distribution, reducing the computational cost to $\mathcal{O}(N \log N)$ in the number of particles N, as opposed to the $\mathcal{O}(N^2)$ required for directly computing all two-body interactions.

The Particle Mesh (PM) approach [85] represents another common way to deal with this issue: in this case, the calculation of the gravitational force is performed by painting the particles onto a 3D mesh of density values ρ using some Mass Assignment Scheme (MAS) (see, e.g., Appendix A), which the potential is then solved for. The force applied to each particle is then obtained by interpolating the force (or potential) computed on the mesh back to the particle's exact position, typically using the same interpolation kernel originally used to assign its mass to the mesh. The computational advantage of this method lies in the fact that the Poisson equation for the gravitational potential

$$\nabla^2 \Phi(\boldsymbol{r}) = 4\pi G \rho(\boldsymbol{r}) \tag{1.34}$$

is easily solved in Fourier space, where it reads

$$\tilde{\Phi}(\boldsymbol{k}) = 4\pi G k^{-2} \tilde{\rho}(\boldsymbol{k}) \quad , \tag{1.35}$$

reducing the complexity to $\mathcal{O}(M \log M)$ where M is the mesh size, thanks to the usage of Fast Fourier Transform (FFT) algorithms [86][3]. This, however, has the drawback of essentially reducing the spatial resolution of the particles to that of the mesh when computing forces, introducing errors in the short range in particular. To solve gravity, the state-of-the-art codes mentioned above implement a combination of these techniques, called TreePM, to leverage the strength of both approaches.

Nonetheless, a large number of simulation time steps is still required to minimize the errors introduced by numerically integrating equations of motion. Vanilla PM solvers are known to fail to reproduce the linear theory growth (the Zel'dovich approximation introduced in Section 1.2.1) at large scale as the number of time steps is reduced (see e.g. [87, 88]). The COmoving-LAgrangian enhanced particle mesh (COLA) scheme [87] proposes to alleviate this issue by computing the large scale displacement using the analytic solution from second order Lagrangian theory (2LPT), which will be discussed below, and the particle mesh is used only to solve for the "residual" small scale displacement that affects halo formation, only adding a relatively small overhead to the vanilla PM method, yet leading to correct growth on large scales. FastPM [89] addresses the same failure mode, yet through a different route. Rather than explicitly splitting the displacement into an analytic 2LPT part and a residual evolved by the PM solver, FastPM modifies the kick and drift operators of the standard leapfrog time integration itself. In a vanilla PM scheme, these operators implicitly assume that the force (and velocity) remain constant over the course of a time step; this assumption is violated even in the linear regime once the steps become coarse, which is precisely the source of the growth errors discussed above. FastPM instead derives its kick and drift factors directly from the Zel'dovich equation of motion, so that the discrete integration reproduces the correct linear growth exactly, regardless of the time-step size. Since this correction is built directly into the integrator, FastPM does not need to separately compute, store, or subtract an LPT displacement field at every step, unlike COLA; this reduces both the code complexity and the memory overhead of the scheme. Additionally, at fixed number of time steps and force resolution, this approach has been found to reduce halo stochasticity relative to COLA [89]. Naturally, the drastic decrease in computational time that these codes achieve by reducing the number of time steps comes with the cost of losing accuracy at small scales (see, for instance, Figure 2.8). Recovering some of this accuracy without reinstating the cost has recently motivated a range of differentiable and data-driven approaches, spanning a wide

[3] The complexity of the MAS itself, both to interpolate the particles' masses to the mesh and the potential on the mesh back to the particles, is linear in the number of particles.

range of model complexity: from low-dimensional, physically motivated corrections fitted by gradient descent [90], through neural corrections acting within the solver itself [91], to deep networks emulating the displacement field outright [92, 93].

All these particle-based methods, whether based on direct summation, hierarchical algorithms, PM, or a TreePM hybrid, are Lagrangian in nature: the quantity being evolved is the trajectory of a fixed set of discrete mass tracers, rather than a density field defined on a static grid, which is the distinguishing feature of an Eulerian approach (the mesh used by PM to solve for the force is merely a numerical device, rebuilt at every step from the particles' current positions). This description can be formalized by labeling each mass element (or particle) by its initial, unperturbed comoving position $\boldsymbol{q}$. As gravitational instability proceeds, this element is displaced to some later position $\boldsymbol{x}(t)$, related to $\boldsymbol{q}$ by

$$\boldsymbol{x}(\boldsymbol{q},\, t) = \boldsymbol{q} + \boldsymbol{\Psi}(\boldsymbol{q},\, t) \quad , \tag{1.36}$$

where $\Psi(\boldsymbol{q},\, t)$ is the Lagrangian displacement field, encoding the full trajectory of each mass element relative to its starting point. Another method that allows to speed up cosmological simulation, especially in the context of large-scale structure formation, is the so-called Lagrangian perturbation theory (LPT) hinted at above [94, 95], which exploits this same formalism to construct an approximate, analytic solution for $\boldsymbol{\Psi}(\boldsymbol{q},\, t)$, in contrast to the full, non-perturbative numerical integration used by N-body simulations, by expanding the initial potential (as opposed to expanding the density field $\rho(\boldsymbol{x},\, t)$ itself, as done in Eulerian perturbation theory). At first order, $\boldsymbol{\Psi}^{(1)}$ reduces to (minus) the gradient of the linearly extrapolated velocity potential, recovering the Zel'dovich approximation introduced in Section 1.2.1; higher orders (2LPT, 3LPT, ...) [96, 97, 98] systematically correct this leading-order solution with source terms built from products of derivatives of the lower-order potentials. Besides serving as an analytic model of structure formation valid on large scales and at early times, LPT is the standard tool used to generate the initial conditions of N-body simulations, by displacing particles from a regular lattice according to $\boldsymbol{\Psi}$ computed from the desired initial power spectrum; using at least 2LPT rather than the Zel'dovich approximation alone is known to substantially suppress the spurious transients that would otherwise contaminate the growth of structure at early times [99]. Crucially, this perturbative description holds only before the trajectories of neighboring mass elements cross (shell-crossing): beyond that point, the expansion of $\boldsymbol{\Psi}(\boldsymbol{q},\, t)$ breaks down, so that only a full, non-perturbative integration of the equations of motion can correctly capture the resulting multi-streaming, virialized dynamics.

1.4.2 From dark matter to observables

N-body simulations achieve their computational advantage by discarding baryons; however, practically every quantity accessible to a survey, e.g. galaxy positions, stellar light, 21-cm emission, Lyα absorption, thermal and kinetic Sunyaev-Zel'dovich signals, is baryonic. Any DM-only approach therefore requires a prescription to reinstate baryonic information in post-processing. Three broad families of methods exist, distinguished by what specifically they predict.

A first family of approaches does not attempt to produce a realization of the tracer field, but rather a prediction for its statistics. Since the tracer density is a local functional of the matter field and must respect its symmetries, it may be expanded in the operators that can be built from the latter,

$$\delta_{\text{tracer}} = b_1\delta + \frac{b_2}{2}\delta^2 + b_{K^2}K^2 + b_{\nabla^2\delta}\nabla^2\delta + \cdots + \varepsilon \quad , \tag{1.37}$$

where K^2 denotes the leading tidal (shear) operator, ε a stochastic contribution, and the coefficients b_i are free parameters to be marginalized over rather than predicted [49, 100, 101, 102]. Equation (1.31) is the leading term of this expansion. Although the expansion is written for the field, what it delivers are correlation functions: these follow as expressions in the b_i and in the statistics of the matter field, so that the scheme provides a parametrized model for $P_{\text{tracer}}(k)$, for its higher-order counterparts, and for their dependence on cosmology, with the stochastic contribution entering only through its own statistics rather than as a particular field. Such perturbative bias expansions, in their Eulerian and Lagrangian formulations, underpin the analysis pipelines of current galaxy surveys, and analogous parametrizations exist for the specific tracers considered in this work: the HI bias and the $M_{\text{HI}}(M_{\text{halo}})$ relation of the HI halo model [103], and the Lyα forest bias parameters together with their non-linear corrections [104]. Hybrid schemes go one step further by combining the expansion with a simulation, weighting N-body particles by functions of their initial Lagrangian density field and displacing them to their evolved positions, so that the non-linear displacement is captured numerically while bias remains perturbative [105, 106, 107]. The limitation of these methods, for the purposes of this thesis, therefore differs between the two cases. The analytic expansions deliver summary statistics with nuisance parameters attached rather than realizations of a field, which is entirely sufficient, and close to optimal, for a likelihood-based analysis of two- or three-point functions, but cannot support inference performed directly at the field level. The hybrid schemes do produce a field, and are in that respect the closest existing counterparts to the approach developed in this thesis, but bias remains perturbative in them and they are correspondingly restricted to scales well above those of interest here, while requiring a full N-body simulation as input.

The second family populates halos with galaxies. The Halo Occupation Distribution (HOD) [108, 109] assigns central and satellite galaxies statistically as a function of halo mass alone, while the Conditional Luminosity Function (CLF) [110] generalizes this to the full luminosity-dependent occupation. (Sub)Halo Abundance Matching (SHAM) [111, 112, 113, 114, 115] instead assumes a monotonic relation between an observable (e.g. stellar mass or luminosity) and a halo proxy (e.g. mass or the maximum circular velocity $V_{\max}$), and matches the two by rank order. More physically motivated semi-analytic models (SAMs) [116, 117, 118, 119, 120, 121, 122] apply simplified prescriptions for gas cooling, star formation and feedback directly to the halo merger trees extracted from the DM-only run, whereas empirical models such as UniverseMachine [123] fit flexible data-driven relations, e.g. between star formation rate, halo mass and halo growth, directly to observations. Cluster observables can be assigned analogously, through calibrated mass–observable scaling relations such as M–L_X or M–Y_{SZ} [124].

The third family predicts fields rather than objects. Baryonification, or Baryon Correction Model (BCM), techniques [125, 126] displace DM particles, or equivalently modify halo density profiles, so as to mimic the redistribution of matter caused by gas ejection and adiabatic contraction, calibrated against full-hydrodynamical simulations; this is the standard route to modeling baryonic effects on the total matter power spectrum, which is essential for weak-lensing analyses. For the diffuse intergalactic medium, gas properties can be painted onto DM-only outputs using analytic halo-based density or pressure profiles [127, 128], from which thermal and kinetic Sunyaev-Zel'dovich or X-ray signals follow, or through the Fluctuating Gunn-Peterson Approximation (FGPA) [70, 129], which assumes the gas density traces the underlying DM density through a power law set by photoionization equilibrium, allowing Lyα forest absorption to be modeled without solving for the hydrodynamics.

These approaches differ widely in physical detail, but the halo-based ones share a structural requirement that matters considerably for what follows: they operate on a halo catalogue, and therefore presuppose that halos are resolved well enough to be identified and to have meaningful masses or velocities assigned to them. This is a strong assumption in the present context: the approximate solvers described in Section 1.4.1 gain their speed precisely by degrading small-scale accuracy, which is where halo properties are determined; halo abundances and internal structure are among the first quantities such solvers get wrong [88, 90, 130]. Methods that act directly on the particle distribution or on the density field, without an intervening halo-finding step, are thus better matched to cheap approximate simulations, motivating the approach developed in this thesis.

In recent years, machine learning (ML)-based methods to connect N-body simulations to baryons have also flourished. Early work employed relatively standard regression tools, e.g. support vector machines, k-nearest neighbors and tree-based methods, to predict

galaxy or gas properties from halo and environmental features [131, 132, 133, 134]. A large body of subsequent work instead treats the task as image-to-image translation, mapping a DM field directly onto a baryonic one through Gaussian processes (GPs) [135] or with convolutional neural networks [136, 137, 138], or as conditional generation, using variational autoencoders and generative adversarial networks (GANs) [139, 140, 141, 142], diffusion models [143, 144, 145] and stochastic interpolants [146]. A further class of methods ignores the dark matter field altogether, generating the baryonic observables directly from the cosmological parameters by using generative models [147, 148]. Such models are expressive enough to reproduce baryonic fields at the field level, and, once trained, are orders of magnitude faster than the full-hydrodynamical simulations they replace; their cost is a large training set of paired DM and full-hydrodynamical runs, limited interpretability, and no guarantee that physical constraints (mass conservation, translational and rotational invariance) are respected other than approximately, as learned from the data.

A model of the kind described in this section, whether physically motivated or purely data-driven, is what will be referred to throughout this thesis as an *emulator*: a surrogate that reproduces the output of an expensive simulation at a small fraction of its cost, ideally as a function of the cosmological and astrophysical parameters that the simulation itself takes as input. The last property is what makes an emulator useful for inference rather than merely for generating individual mock catalogues, as Chapter 4 discusses, and it is the requirement that Chapter 3 examines in detail.

The following chapter introduces `JERALD`, which pursues the same goal from a physics-informed standpoint rather than a purely data-driven direction: it acts on particles instead of fields, requires no halo identification, and enforces mass conservation and the symmetries of the problem by construction, producing output DM or baryonic fields.

Chapter 2

JERALD: a physics-informed tool for accelerating cosmological simulations

The problem of improving approximate N-body cosmological simulations and/or populating DM-only simulations with baryons can be approached from many different points of view. In a machine learning context, as discussed above, substantial effort has been dedicated towards employing state-of-the-art tools to this end; however, a physics-informed perspective could also be adopted.

In general, problems with relatively simple large-scale physics but complex small-scale effects can be tackled within an effective field theory framework. At a high level, the idea is to average over the behavior of the underlying theory at small scales, such that the resulting problem can be approached through a simpler model. Practically, this method consists in rewriting the Lagrangian describing the system under investigation such that it explicitly satisfies the symmetries of the problem, but parametrized by a set of free coefficients accounting for non-perturbative small-scale effects. In the context of cosmological simulations, this strategy has a straightforward application for approximate N-body simulations realized via approximate PM solvers such as COLA and FastPM: as discussed above, the computational gain brought by these solvers comes with the cost of losing accuracy at small scales, due to the way gravity is solved for, the coarse-graining in time and the lack of baryonic feedback; adopting a Lagrangian framework as in Eq. (1.36), one can then compensate for these effects by adding an extra displacement, given by the gradient of a parametrized effective potential, to the DM particles' output position, in order to mimic the missing sub-grid physics [90]. This has the advantage of preserving the symmetries of the problem — i.e., in a cosmological setting, the translational and rotational symmetries — and, on top of that, simply displacing particles explicitly satisfies mass conservation.

However, this approach is not necessarily limited to DM, rather it can be extended to model baryons as well as their properties: the core assumption is that the velocity

of baryons is approximately equal to that of DM, since it is dominated by large scales where the two trace each other. For this reason, using a Lagrangian perspective, the distribution of DM can be thought of as a large-scale distribution of baryons, while the small-scale baryonic physics, similarly to the DM case, can be modeled at the target redshift by displacing the DM particles, now reinterpreted as baryons, such that their final distribution matches that of a specific baryon type or property as given by some reference full-hydrodynamical simulation. This idea was proposed by Dai and Seljak with the name "Lagrangian Deep Learning" (LDL) [149]. `JERALD` is based on LDL, although it introduces modifications and novel features, such as the modeling of neutral hydrogen, which shall be discussed in the following sections.

The work presented in this chapter has been published in [1], of which the author of this thesis is the first author. The paper introduces `JERALD` and focuses on the improvement of DM distributions and in the prediction of stellar mass and neutral hydrogen fields, assessing its performance at fixed cosmology and astrophysics both against the reference used for training as well as new, unseen full-hydrodynamical maps with different initial condition. The model, the results and the figures reported below are those of that paper; the presentation has been reorganized in order to fit the structure of this thesis and to connect the material to the chapters that follow.

2.1 The model

This section closely follows [149] except for small changes. At the highest level, like LDL, `JERALD` takes as input DM particles positions, moves them via one or more layers of Lagrangian displacement (in the work presented here, mostly 2), and finally produces an output density field via a MAS (Cloud-In-Cell or CIC, see [85] and appendix A), optionally applying a field-level transformation to the output. Given a particle p at position $\boldsymbol{x}_p = (x_p, y_p, z_p)$ in comoving coordinates, a Lagrangian displacement layer acts as

$$\boldsymbol{x}_p \to \boldsymbol{x}_p + \boldsymbol{S}_p \ , \tag{2.1}$$

where the displacement field $\boldsymbol{S}_p$ is modeled as the gradient of an effective potential, the source of which, similarly to gravity, is a simple power law of the normalized density (omitting the dependence on $\boldsymbol{x}$ for ease of notation):

$$\boldsymbol{S}_p = \boldsymbol{\nabla}_p \, \phi(\delta) = \boldsymbol{\nabla}_p \big(\hat{G} f(\delta)\big) \quad , \qquad f(\delta) = (1+\delta)^\gamma \ , \tag{2.2}$$

where $\boldsymbol{\nabla}_p$ is the gradient calculated at particle position p and

$$1 + \delta(\boldsymbol{x}) = \rho(\boldsymbol{x}) / \sum_p \rho(\boldsymbol{x}_p) \tag{2.3}$$

is the normalized density, computed assuming constant DM particle mass (which cancels out when normalizing by the average density). In Eq. (2.2), $\hat{G}$ is the Green's operator, the action of which can be explicitly written as

$$\hat{G} f(\delta)(\boldsymbol{x}) = \int G(\boldsymbol{x}, \boldsymbol{x}') f(\delta(\boldsymbol{x}')) d\boldsymbol{x}' \ , \tag{2.4}$$

with $G(\boldsymbol{x}, \boldsymbol{x}') = G(\boldsymbol{x} - \boldsymbol{x}')$ because of translational symmetry. Being a convolution, Eq. (2.4) has a straightforward expression in Fourier space as

$$\mathcal{F}\big[\hat{G} f(\delta)\big](\boldsymbol{k}) = \tilde{G}(\boldsymbol{k})\, \tilde{f}(\delta)(\boldsymbol{k}) \ , \tag{2.5}$$

with $\tilde{G}(\boldsymbol{k}) = \tilde{G}(k)$ due to rotational symmetry.

Eq. (2.2) features the first free (learnable) parameter, γ, the purpose of which is to balance the contributions of large and small values in the input density field, making it dependent on the specific structures present in the starting (and target) simulation, particularly in the case of small simulation boxes. The next chapter will investigate the behavior of this parameter and this dependence in detail.

The explicit form of $\tilde{G}(k)$ depends on the target quantity: when modeling stellar mass, the Green's operator is parameterized, as in LDL, in the following way:

$$\tilde{G}_{\rm sm}(k) = \alpha_{\rm sm} \exp\left[-\left(\frac{k^h_{\rm sm}}{k}\right)^2\right] \exp\left[-\left(\frac{k}{k^l_{\rm sm}}\right)^2\right] k^{\nu_{\rm sm}} \ , \tag{2.6}$$

On the contrary, in the case of DM, we employ an expression for the Green's function which is slightly different with respect to LDL, as we find it works best:

$$\tilde{G}_{\rm dm}(k) = \exp\left[-\frac{k^h_{\rm dm}}{k}\right] \exp\left[-\frac{k}{k^l_{\rm dm}}\right] k^{\nu_{\rm dm}} \ . \tag{2.7}$$

In these expressions, α, k^h, k^l and ν are again free (learnable) parameters. The purpose of the α parameters is to calibrate the strength of the correction, the first exponential factor suppresses large-scale growth for $k < k^h$, since the physics to model through an effective description happens at small scales, while the second term mimics force softening. The final power law term is motivated by the fact that fixing $\gamma = 1$ in Eq. (2.2) and setting $\alpha = 4\pi G\bar{\rho}$, $k^h = 0$, $k^l = \infty$ and $\nu = -2$ in Eq. (2.7) or Eq. (2.6) retrieves the gravitational potential as in Eq. (1.35). In principle, all free parameters may depend on redshift, the

level of accuracy of the starting simulation, the structures present in it, and the specific implementation of baryonic interactions in the target simulation; additionally, especially in the case of Eq. (2.7), parameters $\alpha_{\rm dm}$, $k^l_{\rm dm}$ and $\nu_{\rm dm}$ may be degenerate over large ranges of wavevectors: this will be discussed more in detail in the next chapter.

In practice, each displacement layer is implemented as follows. Given particle positions $\boldsymbol{x}$, a density field is produced via CIC and raised to a power (learned and different between displacement layers), as indicated in Eq. (2.2). The potential in Fourier space is then computed as the product of the Discrete Fourier Transform (DFT) of this field and the Green's function parameterized based on the quantity to model—i.e. as in Eq. (2.7) for DM, or as in (2.6) for stellar mass. The gradient in Eq. (2.2) is computed directly in Fourier space as the calculation is straightforward, since, in real space, a finite difference approximation of a derivative can be written as a convolution, and so it simply translates into a multiplication by a filter in Fourier space (see Appendix D for considerations on this). Finally, the gradient of the potential is transformed back into real space and interpolated from the mesh points of the CIC mesh to the particle positions $\boldsymbol{x}$, using an interpolation scheme analogous to CIC.

After the displacement layers, a density map $\rho_{\rm out}$ is produced again using CIC. The final transformation applied to $\rho_{\rm out}$ depends on the target map type and is designed to capture the physical processes that cannot be modeled simply via matter transport, such as star formation for a stellar mass target and ionization for neutral hydrogen.

For stellar mass, the transformation has the form:

$$F_{\rm sm}(\boldsymbol{x}) = \mathrm{ReLU}(w_{\rm sm}(1+\delta_{\rm out}(\boldsymbol{x}))^{\mu_{\rm sm}} - b_{\rm sm}) \ , \tag{2.8}$$

where $w_{\rm sm}$, $\mu_{\rm sm}$ and $b_{\rm sm}$ are free (learnable) parameters and $\delta_{\rm out}$ is the overdensity associated to the density $\rho_{\rm out}$. This transformation is meant to select regions of high density where star formation occurs.

In the case of DM, no final transformation is applied, implying

$$F_{\rm dm}(\boldsymbol{x}) = 1 + \delta_{\rm out}(\boldsymbol{x}) \ . \tag{2.9}$$

The free parameters of the model are optimized by matching the output field $F(\boldsymbol{x})$ to a distribution of the physical quantity to be modeled $F_{\rm ref}(\boldsymbol{x})$, obtained from a high-resolution full-hydrodynamical reference simulation. This is achieved by minimizing an L1 or L2 loss:

$$\mathcal{L} = \frac{1}{N_{\rm mesh}} \left\| \hat{O}_s F(\boldsymbol{x}) - \hat{O}_s F_{\rm ref}(\boldsymbol{x}) \right\|^p \ , \tag{2.10}$$

where $\|\cdot\|^p$ is the p-th norm ($p = 1, 2$), $N_{\rm mesh}$ is the number of mesh points (or voxels) in the maps F, $F_{\rm ref}$ and $\hat{O}_s$ is an optional smoothing operator (used for the stellar mass

maps and HI at $z = 5$ only), defined in Fourier space as

$$\tilde{O}_s(k) = 1 + \left(\frac{k}{1h\mathrm{Mpc}^{-1}} \right)^{-n} , \tag{2.11}$$

where n is a hyperparameter that determines the relative weight between the large-scale and small-scale modes, introduced to prevent the model from focusing predominantly on the small scales. Note that, importantly, for a sensible voxel-based comparison, the target full-hydrodynamical and the approximate N-body simulations need to share the same initial white noise map. The way this is achieved in this work is detailed below. Similarly to the original implementation, in the case of DM, we use an L2 loss and no smoothing, while for stellar mass we use an L1 loss and a different n for different redshifts, as we shall explain below. A schematic of the full pipeline for DM and stellar mass is illustrated in Figures 2.1 and 2.2.

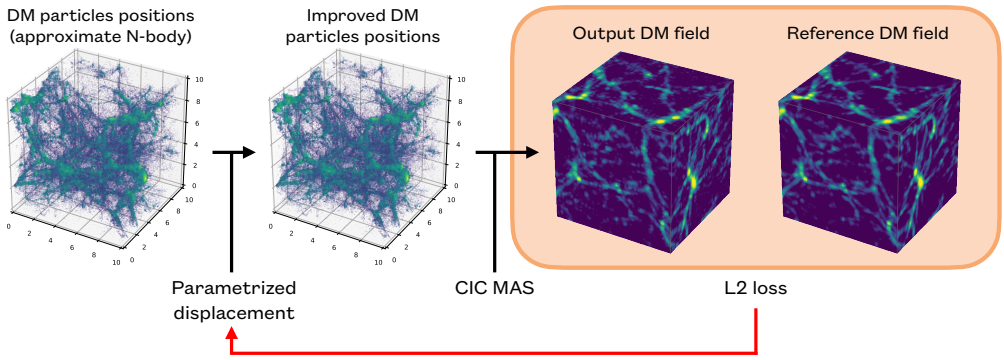


Figure 2.1: Visual pipeline for how `JERALD` improves an input DM distribution. The orange box highlights what the L2 loss is calculated over, and the red arrow indicates which parameters are updated via the gradient of the loss.

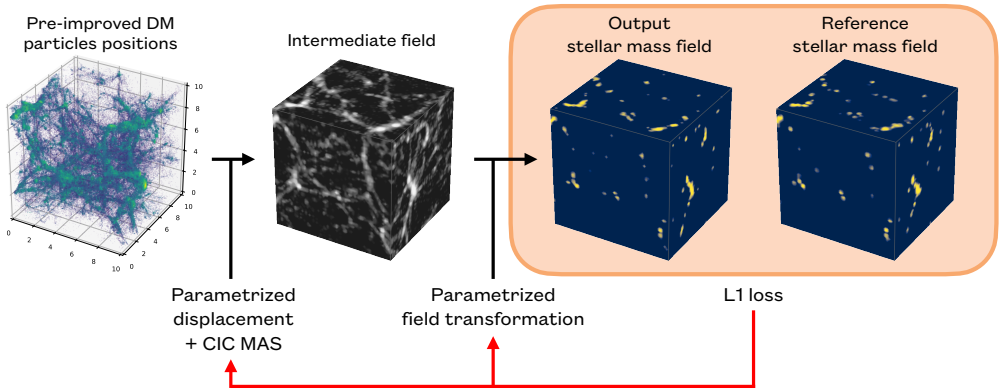


Figure 2.2: Same as Figure 2.1 but for stellar mass. The field produced after the displacement layers is an intermediate step generated before the field transformation and cannot be interpreted as any physical field. The stellar mass prediction starts from particle positions already improved by a `JERALD` pre-trained on the DM field.

2.2 Neutral hydrogen

Going beyond [149], we now extend the model to predict HI maps. In this case, the key challenge is represented by galactic feedback, a complex process whose impact on the HI distribution propagates from small scales, where it originates, to large scales. The role of feedback on the HI distribution is similar to that on the overall gas distribution: at small/intermediate scales, galactic winds are likely to push gas outside galaxies in the low-density intergalactic medium (IGM), either with thermal or mechanical feedback, while at small scales baryons can condense and determine an increase of power compared to a simulation in which baryons are absent (see e.g. [150]). HI is expected to follow, at least to some extent, the gas distribution, providing an overall suppression of power at small/intermediate scales and increase of power at very small scales [151, 152, 153]; however, differently from other baryonic components, the HI distribution is also sensitive to the local gas temperature, UV background and radiative transfer effects from astrophysical sources that ionize the surrounding medium, including patchy reionization [78].

Furthermore, the dynamical range of HI is much greater compared to that of, e.g., stellar mass, and it can stretch over several orders of magnitude, from neutral fractions of 10^{-7} in filaments up to fully neutral gas inside galaxies in self-shielded regions at low redshift [154]: for this reason, devising an approximate model that can reproduce the HI distribution in its entire dynamical range is a very challenging task. Nonetheless, different parts of this range can be probed through different observables: in particular, the high-density regions, found within halos (i.e., at $z = 0$, with number densities above $\sim 10^{-10}\mathrm{cm}^{-3}$ and up to $\sim 10^{-5}\mathrm{cm}^{-3}$), are observable via 21cm intensity mapping, while the low to intermediate densities, found in the IGM (at $z = 0$, with HI number density reaching down to $\sim 10^{-17}\mathrm{cm}^{-3}$ at $z = 0$), are relevant for the study of the Lyα forest.

In this work, we focus on the IGM; clearly, using the loss in Eq. (2.10) to train `JERALD` on a reference simulation that models the full dynamical range of HI without any pre-processing would result in the model focusing solely on the high-density range, which would dominate the loss. Since, for the purpose of modeling the Lyα forest, we are not interested in the high-density regions, this problem can be circumvented, for example, by clipping the HI density above a given threshold in the reference simulation, thus allowing the cosmic web to stand out and be learned. Here, this step is not necessary, as the reference simulations utilized do not implement self-shielding [77]. This results in the HI inside halos being greatly underestimated: this is an unrealistic scenario, but if one focuses on the IGM rather than the halos, implementing self-shielding is not strictly necessary, and this is actually beneficial for our purpose as it reduces the dynamical range of HI. More generally, if the main focus is the IGM, the features of the distribution one has to

model depend on the prescription used to mask out the high-density regions: the approach described below is therefore well-suited for reference simulations without self-shielding, while, for example, clipped maps would require a different treatment—we plan to explore this in the future.

In practice, to model neutral hydrogen, we modify the original LDL approach in two ways. Firstly, we introduce a new term in the effective potential:

$$\boldsymbol{S} = \boldsymbol{\nabla}\big(\hat{G}^{(1)}_{\rm HI} f^{(1)}_{\rm HI}(\delta)\big) + \boldsymbol{\nabla}\big(\hat{G}^{(2)}_{\rm HI} f^{(2)}_{\rm HI}(\delta_{\rm star})\big) \ , \tag{2.12}$$

where both Green's operators $\hat{G}^{(1)}_{\rm HI}$ and $\hat{G}^{(2)}_{\rm HI}$ are defined as in Eq. (2.6) and have separate free parameters, $f^{(1)}_{\rm HI}$ and $f^{(2)}_{\rm HI}$ are power laws with separate free exponents as in Eq. (2.2), δ is once again the input particles overdensity while $\delta_{\rm star}$ is the stellar overdensity associated to the distribution δ. The purpose of the second term in Eq. (2.12) is to approximately model the effect of galactic winds. In order to predict HI using solely the input DM particle positions, for $\delta_{\rm star}$ we use the prediction obtained from `JERALD` as described in the previous section and pre-trained separately, i.e. $1 + \delta_{\rm star} = F_{\rm sm}$ from Eq. (2.8).

Secondly, we replace the final transformation of Eq. (2.8) with the expression:

$$F_{\rm HI}(\boldsymbol{x}) = {\rm ReLU}(w_{\rm HI}(1 + \delta_{\rm out}(\boldsymbol{x}))^{\mu_{\rm HI}} - D(\boldsymbol{x}) - b_{\rm HI}) \ , \tag{2.13}$$

where $D(\boldsymbol{x})$ is a depletion term given by

$$D = {\rm ReLU}(\beta^{(1)}\hat{O}_{\rm I}\,(1 + \delta_{\rm star}) + \eta) + \beta^{(2)} f^{(3)}(\delta_{\rm star}) \ , \tag{2.14}$$

where $\beta^{(1)}$, $\beta^{(2)}$ and η are free parameters, $f^{(3)}$ is once again a power law as in Eq. (2.2) with a different free exponent and $\hat{O}_{\rm I}$ is a filter defined in Fourier space as

$$\tilde{O}_{\rm I}(k) = \exp\left[-\frac{k^h_{\rm I}}{k}\right] \exp\left[-\frac{k}{k^l_{\rm I}}\right] k^{\nu_{\rm I}} + \xi \ , \tag{2.15}$$

with $k^h_{\rm I}$, $k^l_{\rm I}$, $\nu_{\rm I}$ and ξ learnable parameters. Heuristically, the first term in Eq. (2.14), which affects the HI distribution both where stars are present and in their surroundings, is meant to capture the depletion of HI due to processes such as ionization, while the second is more local and it accounts for the conversion between gas and stars. We train all parameters as discussed in the previous section, using an L1 loss (see Eq. (2.10)) as we find it produces the best results.

The pipeline of the `JERALD` approach to produce HI maps is illustrated in Figure 2.3. All free parameters in the model for DM, stellar mass and HI, are summarized in Table 2.1.

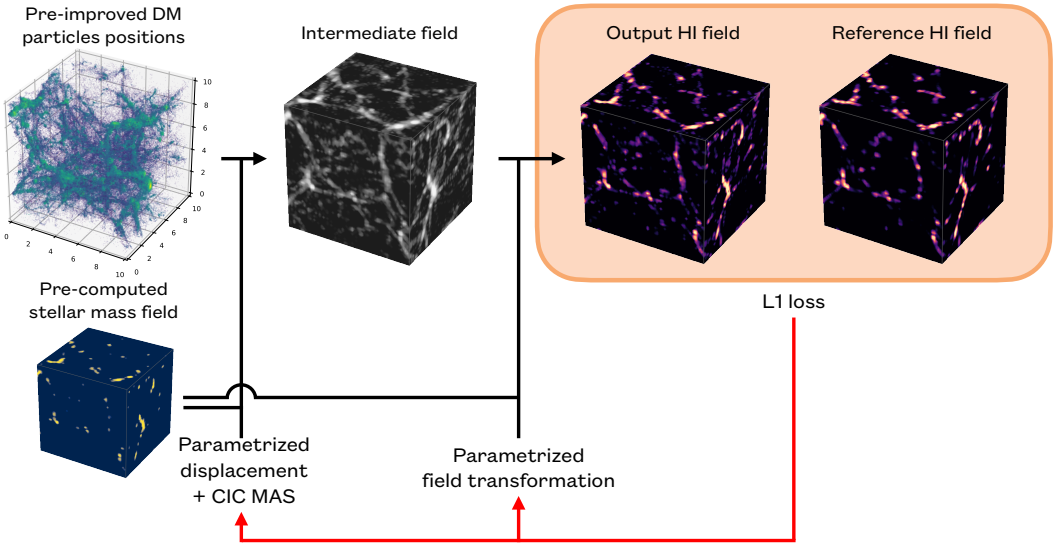


Figure 2.3: Same as Figure 2.1 but for neutral hydrogen. The intermediate field generated before the field transformation is again colored in gray for the same reason discussed in Figure 2.2. The figure also highlights the presence of a stellar mass field, pre-computed with `JERALD`, both in the displacement(s) (Eq. (2.12)) and in the field transformation (Eq. (2.13)).

Table 2.1: `JERALD` parameters for DM, stellar mass and HI targets, at a given redshift. N_p is the number of parameters in each part of the model and $N_p^{\rm tot}$ the total number of parameters for each predicted quantity, with $S_{\rm dm}$, $S_{\rm sm}$ and $S_{\rm HI}$ being the number of displacement layers used for DM, stellar mass and HI respectively (all set to 2 except for $S_{\rm sm} = 3$ at $z = 5$). $N_p^{\rm used}$ is the total the number of parameters used in this work (except for stellar mass at $z = 5$, where $N_p^{\rm used} = 18$ (+10)). For stellar mass and HI, the total number of parameters also includes, in brackets, those used to improve the DM distribution (for both) and to produce the stellar mass map (for HI).

Quantity		Equation	Parameters	N_p	$N_p^{\rm tot}$	$N_p^{\rm used}$
Dark matter	Displacement	(2.2)	$\gamma_{\rm dm}$	$S_{\rm dm}$	$5S_{\rm dm}$	10
	Green's function	(2.7)	$\alpha_{\rm dm}$, $k^h_{\rm dm}$, $k^l_{\rm dm}$, $\nu_{\rm dm}$	$4S_{\rm dm}$		
	Transformation	-	-	-		
Stellar mass	Displacement	(2.2)	$\gamma_{\rm sm}$	$S_{\rm sm}$	$5S_{\rm sm}+3$ ($+5S_{\rm dm}$)	13 (+10)
	Green's function	(2.6)	$\alpha_{\rm sm}$, $k^h_{\rm sm}$, $k^l_{\rm sm}$, $\nu_{\rm sm}$	$4S_{\rm sm}$		
	Transformation	(2.8)	$w_{\rm sm}$, $\mu_{\rm sm}$, $b_{\rm sm}$	3		
Neutral hydrogen	Displacement	(2.12)	$\gamma^{(1)}_{\rm HI}$	$S_{\rm HI}$	$10S_{\rm HI}+11$ ($+5S_{\rm dm}$ $+5S_{\rm sm}+3$)	31 (+10 + 13)
			$\gamma^{(2)}_{\rm HI}$	$S_{\rm HI}$		
	Green's function(s)	(2.6)	$\alpha^{(1)}_{\rm HI}$, $k^{h\,(1)}_{\rm HI}$, $k^{l\,(1)}_{\rm HI}$, $\nu^{(1)}_{\rm HI}$	$4S_{\rm HI}$		
			$\alpha^{(2)}_{\rm HI}$, $k^{h\,(2)}_{\rm HI}$, $k^{l\,(2)}_{\rm HI}$, $\nu^{(2)}_{\rm HI}$	$4S_{\rm HI}$		
	Transformation	(2.13)	$w_{\rm HI}$, $\mu_{\rm HI}$, $b_{\rm HI}$	3		
		(2.14)	$\beta^{(1)}$, $\beta^{(2)}$, η, $\gamma^{(3)}_{\rm HI}$	4		
		(2.15)	$k^h_{\rm I}$, $k^l_{\rm I}$, $\nu_{\rm I}$, ξ	4		

2.3 Results

2.3.1 Code specifications

The `JERALD` code is written partly in Python using `JAX` [155] and partly in `C++`. The CIC paint and readout (see Appendix A) and the DFTs are implemented in `C++`, the latter via MPI-enabled FFTW [156], and bound to Python via Pybind11 [157]. In Python, CIC paint and readout as well as forward and backward DFTs are defined as JAX primitives and equipped with definitions of vector-Jacobian products (see Appendix C), making them jittable and available for automatic differentiation. All methods are parallelized using MPI (see Appendix B)—in Python, via MPI4JAX [158]. Each MPI rank deals with approximately $1/W$ of the particles and $1/W$ of the density fields, where W is the size of the MPI communicator, making the code scale approximately inversely proportional in time and constant in memory as W increases (see Figure 2.4). For the 512^3 particles and 512^3 mesh points maps discussed in this work, each step in the optimization (evaluation of training and validation loss and training loss gradient) takes ∼15-20s per displacement layer in one node of the Leonardo supercomputer in CINECA (32 CPUs). The number of optimization steps depends on the kind of reference map and on the optimizer parameters; generally, at most 500 steps are sufficient for the validation loss to reach convergence, and one training typically takes ∼2-5hrs, depending on the number of displacement layers (in this work, we used at most 3). Evaluation of the maps with a pre-trained model takes ∼10s with the same resources.

For a comparison of speed and memory usage with the original LDL code from [149], in Figure 2.4 we show time per optimization step (left) and resident set size (right) for 20 training steps in the case of a 512^3 voxels DM target with 512^3 particles and one displacement layer (the number of CPUs is also the size of the MPI communicator W). We observe that, while the scaling law for the time per optimization step is approximately the same, the proportionality coefficient for `JERALD` is significantly smaller with respect to the one of LDL; nonetheless, the curves for the RSS indicate that `JERALD` is also more memory efficient than LDL.

We now turn to evaluating `JERALD`'s performance. We first give details about the reference simulations and the training procedure, then we investigate the quality of `JERALD`'s outputs against the same maps used for training, and finally we provide a first assessment of its generalization capabilities by applying the trained model to an approximate N-body simulation with new initial conditions (ICs) and compare the output maps with targets that were not seen during training.

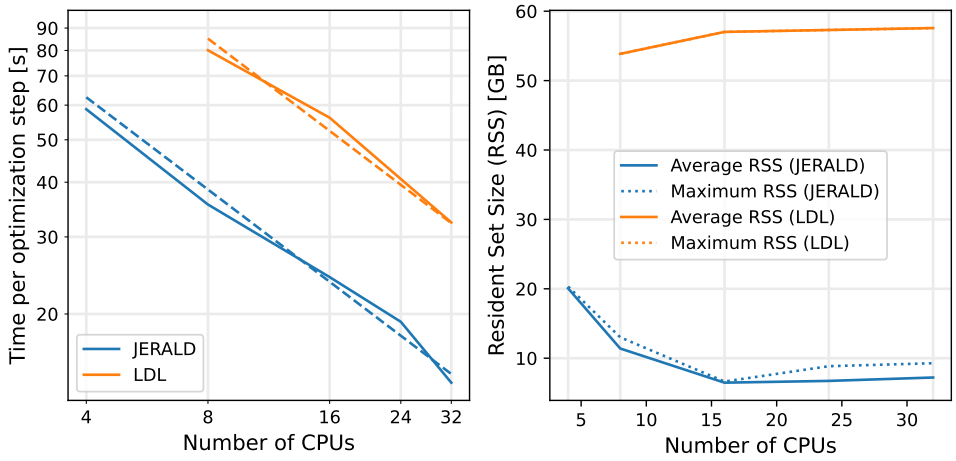


Figure 2.4: Time per optimization step (left) and resident set size (RSS, right, a measure of memory used) for `JERALD` and the original LDL code on one node of the Leonardo supercomputer. The RSS is given by the slurm tool `sacct`, which may not capture rapid memory usage spikes. Times are given as averages over 20 optimization steps in the case of 512^3 particles and a 512^3 voxels DM target map and 1 displacement layer, and the RSS values are those associated with the 20 steps. The dashed lines in the plot on the left indicate the power law $\propto (N_{\rm CPUs})^{-0.7}$ for reference. The maximum RSS for LDL is not visible as it coincides with the average RSS.

2.3.2 Simulations and training

The reference maps used in the subsequent section were extracted from the Sherwood-Relics simulation suite [77, 78]. This set of simulations follows the evolution of DM, gas and stars in a periodic cosmological volume, and the key physical ingredients are: *i*) an evolving UV background that returns a thermal history for the IGM which is in agreement with observations; *ii*) a simplified star formation criterion which converts into star particles all the gas particles whose overdensities are larger than 1000 times the mean baryon density and whose temperatures are below 10^5 K. These simulations are supposed to offer an accurate description of the filamentary cosmic web, and in particular Lyα forest statistics in agreement with observations [159]. The specific simulations used here feature a box size of $80h^{-1}$Mpc and 1024^3 DM and gas particles, with fiducial parameters $\Omega_m = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_b = 0.0482$, $h = 0.678$, $\sigma_8 = 0.829$, and $n_{\rm s} = 0.961$.

Input DM particles positions for `JERALD` are computed using the approximate N-body solver FastPM [89]; we generate the ICs for 512^3 particles at $z = 9$ using 2LPT with a $(2 \times 512)^3$ linear density map, and run 10 FastPM steps up to the desired redshift values ($z = 0, 2, 5$). Crucially, the initial white noise map is generated using a random number generator compatible with the one used for the target Sherwood-Relics simulation (part of the N-GenIC code by [160]) and the same seed, as well as same initial power spectrum and σ_8 (through the normalization constant of the power spectrum). Starting from the FastPM input, `JERALD`'s outputs were computed using the same 512^3 particles and a 512^3

mesh for all fields, making the resolution of the output maps $8\times$ smaller than the reference simulation.

We use the Adam optimizer [161] with a scheduled learning rate ranging from 0.002 to 0.04 after extensive experimentation with other optimizers and hyperparameters. Model parameters are initialized to arbitrary values, and, in the code, they are multiplied by constants to loosely guarantee them to vary in the same range for a more stable training procedure. Specifically, all parameters are expected to be $\mathcal{O}(1)$ except for the coefficients in front of the Green's functions (e.g. α in Eq. (2.6)), which are generally of order 10^{-2}. Yet, parameters that are initially considerably far from the optimal ones may be associated with very large gradients; for this reason, we adopt a time-varying schedule for the learning rate $\eta(i)$, increasing it during training following a softmax schedule:

$$\eta(t) = (\eta(N_{\rm step}) - \eta(0)) \left(\frac{2}{e^{-t/\tau_\eta} + 1} - 1 \right) + \eta(0) \tag{2.16}$$

where generally, for $N_{\rm step}$ training steps, $\eta(0) \sim \mathcal{O}(10^{-3})$, $\eta(N_{\rm step}) \sim \mathcal{O}(10^{-2})$ and $\tau_\eta \sim 100$.

The training/validation/test sets splitting is chosen as follows: for a $M \times M \times M$ map, where M is the number of voxels in each direction, we use a $(0.5M) \times M \times M$ box for training, a $(0.43M) \times (0.43M) \times (0.43M)$ box for testing, and the rest for validation, which correspond to approximately 50% of the pixels in the original map for training, 42% for validation and 8% for testing.

As shown in Table 2.1, all results included in the next sections were obtained using 2 displacement layers for each target at each redshift, except in the case of stellar mass for $z = 5$, where 3 displacement layers were utilized. We adopted a unit n hyperparameter for the smoothing operator in Eq. (2.11) in the case of stellar mass at $z = 0$ and $z = 2$ and HI at $z = 5$, while for stellar mass at $z = 5$ $n = 0.7$ was found to give the best result. In order to bring all training values to the same scale, of order unity, all calculations involving stellar mass were performed in units of $10^{10}h^{-1}M_\odot$, while target HI number densities were divided by a redshift-dependent constant. Note that optimizations starting from different initial guesses for the model parameters can lead to different optimized values, yet with very similar performances. This is discussed in further detail in Section 2.4, and will be analyzed extensively in the next chapter. As this indicates the presence of similar local minima, or non-trivial degeneracies in parameter space, assigning uncertainties to the optimized parameters is not particularly informative, so we refrain from doing so.

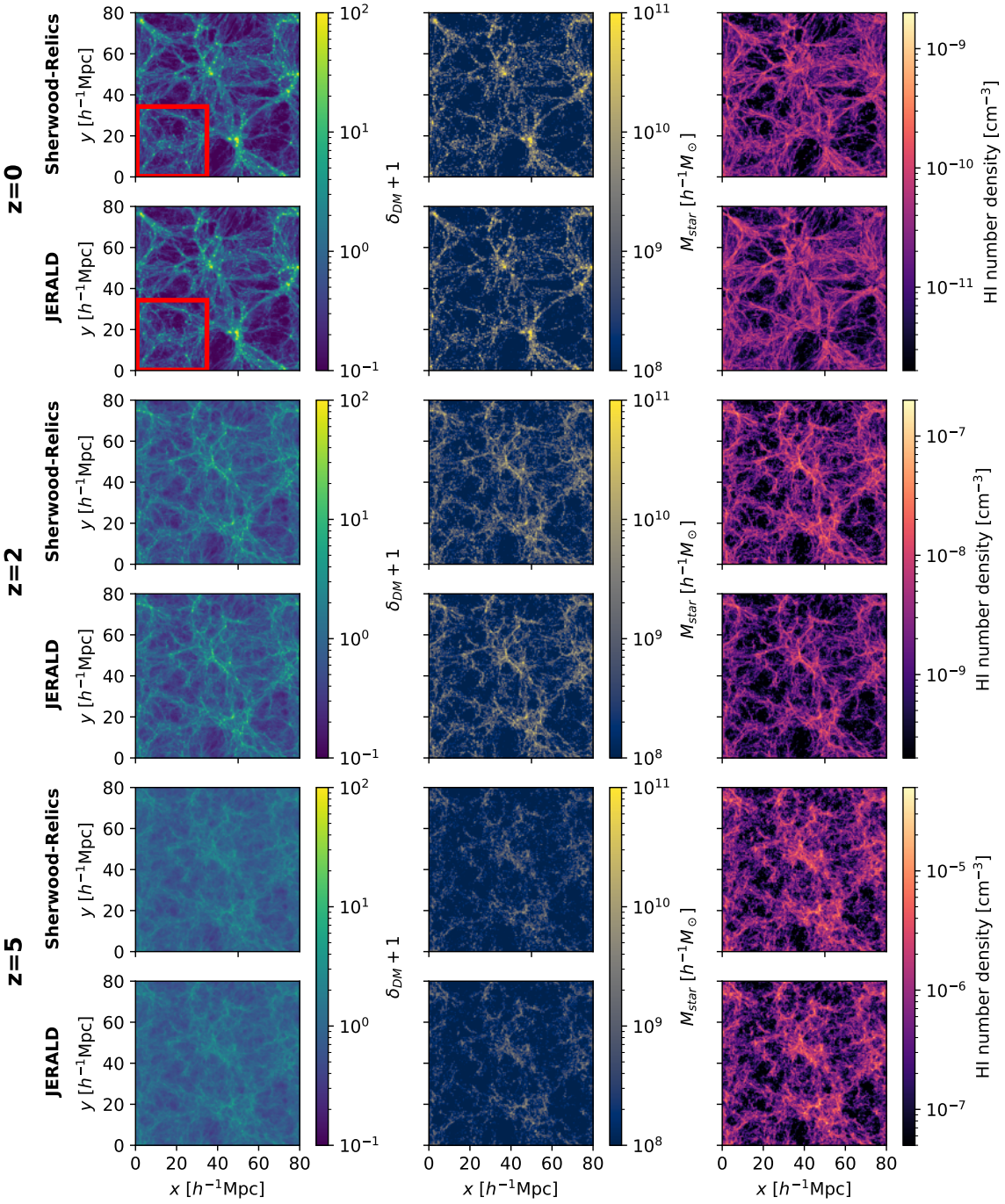


Figure 2.5: Dark matter overdensity, stellar mass and neutral hydrogen number density maps, from left to right. Each pair of maps compares the reference map from Sherwood-Relics (top panel) with `JERALD`'s output (bottom panel), for $z = 5$, $z = 2$ and $z = 0$ (top to bottom). Notice the different range in the colormaps for neutral hydrogen. All images were obtained as $10h^{-1}$Mpc z-slices of the full $(80h^{-1}\text{Mpc})^3$ simulation box. All voxels used to produce the images were part of either the validation or the test set; the red rectangles in the top left pair of maps—omitted in the others—represent the test sub-box, and it extends for the full length of the z-slice used to produce these images. Note that the HI map for $z = 0$ was normalized with the average of the reference map.

2.3.3 Evaluation of JERALD's output maps

Figure 2.5 shows slices of dark matter overdensity, stellar mass and neutral hydrogen number density maps obtained using JERALD from an input approximate DM distribution from FastPM at redshift $z = 0$, $z = 2$ and $z = 5$, compared with the respective Sherwood-Relics references. The red squares shows the test region – i.e., the part of the cube that has not been used for training nor validation. All the voxels used to produce the images in the figure were either part of the validation or the test sets. The visual agreement is remarkable.

To evaluate more quantitatively the performances of JERALD, in Figures 2.6 and 2.7 we also show the relative difference between the reference maps and JERALD's outputs, both post-processed in two steps[1]. Firstly, each map was clipped below $0.5\,\bar{\phi}_i$, where $\bar{\phi}_i$, $i = \mathrm{dm/sm/HI}$, are the averages of the reference Sherwood-Relics DM, stellar mass or HI maps. This is because we expect mesh points with small density/mass to only bring minor contributions to the quantities we plan to compute using maps produced by JERALD. By construction, the loss in Eq. (2.10) prioritizes matching mesh points with large values; for this reason, mesh points with small values, being very abundant, may produce large relative differences which would dominate over the ones associated to the actual range of interest. Indeed, in each case (DM, stellar mass and HI), the clipped values are above 80% of the entire map, but the mass contribution of the rest of the map is at least 10 times greater. The clipping threshold is chosen arbitrarily and it relies on the knowledge of the averages of the reference maps $\bar{\phi}_i$, $i = \mathrm{dm/sm/HI}$, only to define a value that is small *relative to each map type*; of course, when using JERALD on unseen maps, these averages would not be available nor necessary. Secondly, each map was smoothed using a Gaussian filter with a radius of $200h^{-1}$kpc, since, in view of using this approach to model real data, we expect very small scale differences not to lead to significant mismatches. We only illustrate the relative differences for $z = 0$, as distributions become progressively more nonlinear for lower redshifts and therefore more complicated to model—the relative difference maps for $z = 2$ and $z = 5$ display considerably less features than the ones illustrated here. Despite the high resolution and small smoothing radius, the relative differences are characterized by only several notable features, with most of the maps lying below 20%. Note that the locations of the prominent features present in the maps are approximately the same.

[1] We emphasize that these two steps were only implemented when computing relative differences; for all plots of each individual map, all power spectra and all cross correlation coefficients the maps used are the ones produced by JERALD, unprocessed.

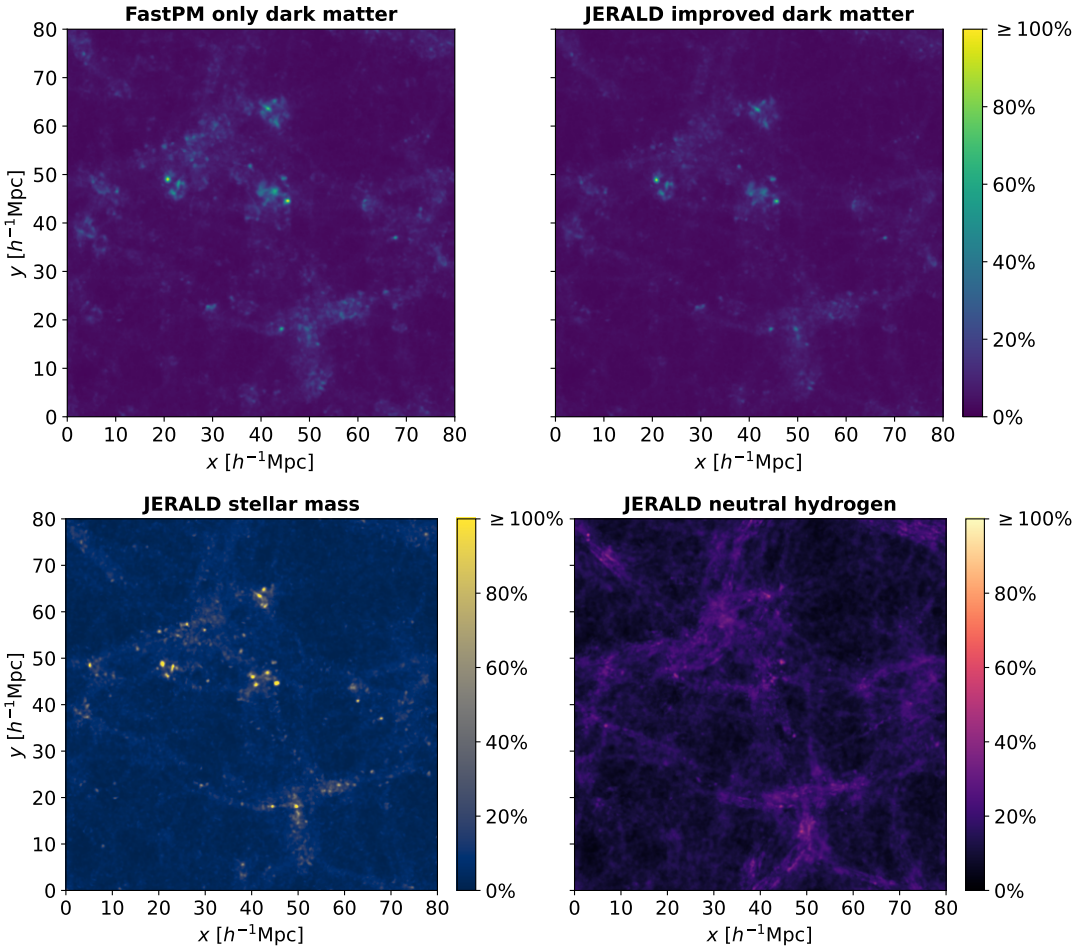


Figure 2.6: Relative difference between $z = 0$ FastPM only and JERALD's output for dark matter, stellar and neutral hydrogen maps. Each map used to compute the relative differences was clipped below 0.5 times the average mass of the reference map and smoothed with a Gaussian filter with a radius of $200h^{-1}$kpc. The plots show the relative difference averaged over the entire length of the $80h^{-1}$Mpc box along the z direction.

To analyze the statistics of JERALD's outputs, we also compare the power spectra of the maps with the reference ones in Figures 2.8 and 2.9 for dark matter, stellar and neutral hydrogen overdensities respectively. Similarly to [149], the power spectra were computed on the entire $(80h^{-1}\text{Mpc})^3$ simulation box, rather than on the test subset only, in order to capture the large scales as well and determine whether the model affects the distributions at those scales. Including the training data in this analysis may bias the results, therefore in the next section we will also analyze the performance of the approach on unseen maps. The bottom panels of Figures 2.8 and 2.9 show the cross correlation coefficient, calculated as:

$$c(k) = \frac{P_{\text{cross}}(k)}{\sqrt{P_{\text{ref}}(k) P_{\texttt{JERALD}}(k)}} \tag{2.17}$$

where $P_{\text{cross}}(k)$ is the cross power spectrum and $P_{\text{ref}}(k)$ and $P_{\texttt{JERALD}}(k)$ are the power spectra of the Sherwood-Relics reference simulation and JERALD, respectively. All power

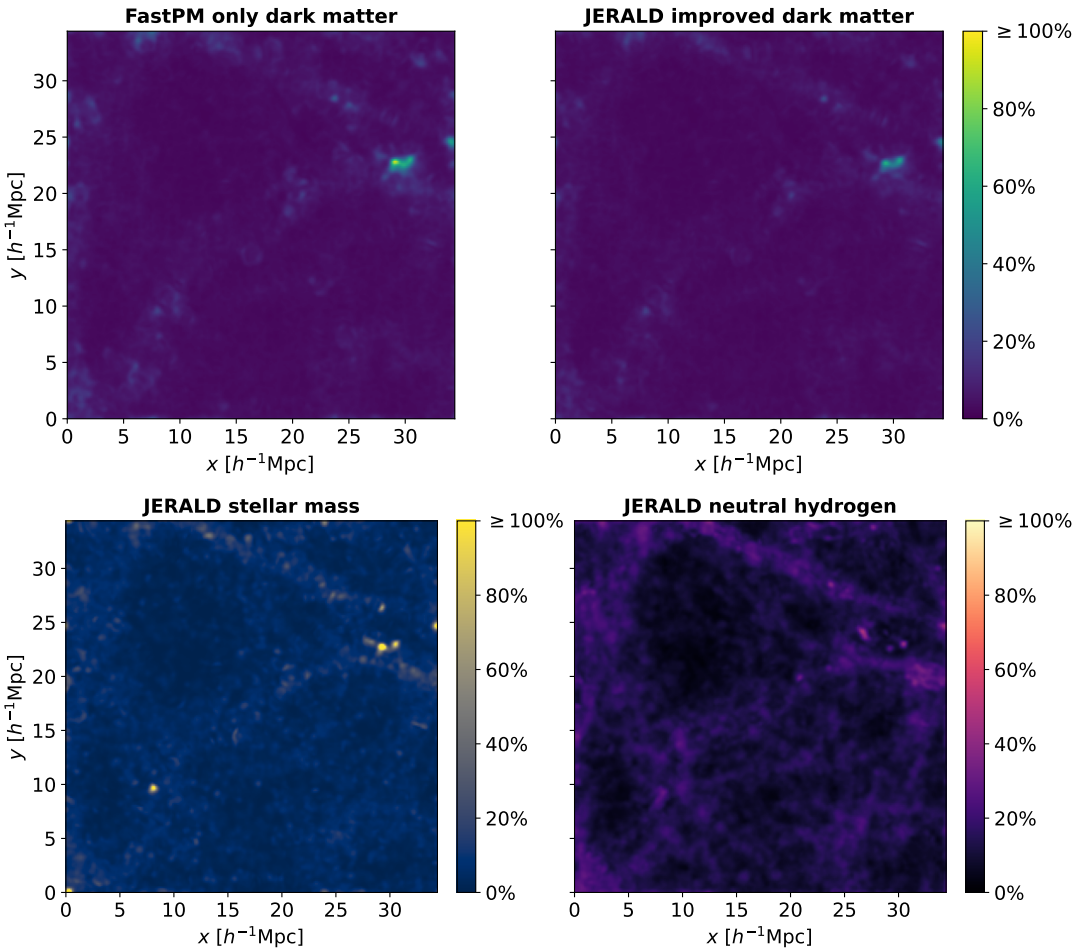


Figure 2.7: As in Fig. 2.6, but showing the relative difference only for the test set, averaged over the $34.4h^{-1}$Mpc of the test sub-box along the z direction.

spectra and cross correlation coefficients were calculated using the Pylians libraries [162]. In each case and at each redshift, the power spectra of the maps produced by `JERALD` are in excellent agreement with the reference ones even up to highly non-linear scales. As expected, the cross correlation coefficients show that the agreement between the maps increases as redshift increases. Note that, in computing the stellar overdensities and HI overdensity at $z = 5$, we used the average of the reference maps rather than that of the maps produced by `JERALD`. This was done because the introduction of the smoothing filter of Eq. (2.11) in the loss leads to sharp mass peaks being smeared out and, as a consequence, the average of the map produced by `JERALD` is smaller than the one of the reference (though only by a small factor of $\sim 5\%$). Nonetheless, these mass peaks do not bring significant contributions to the overall shape of the power spectra, as the left panel of Figure 2.8 and Figure 2.9 highlight, and we expect them not to pose problems in applications of this method.

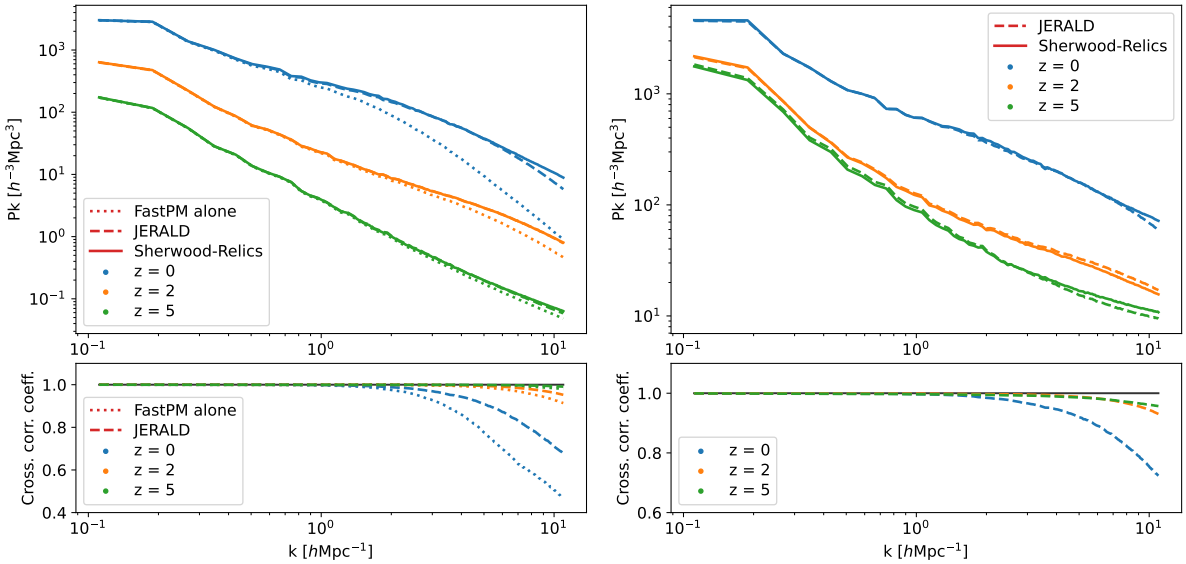


Figure 2.8: *Left:* DM power spectra (top) and cross correlation coefficients (bottom) for the reference Sherwood-Relics simulation (solid), the approximate N-body FastPM simulation (dotted) and the map improved by `JERALD` (dashed), for three different redshifts. For $z = 2$ and $z = 5$, the improved power spectra are almost not visible as they overlap with the reference ones. *Right:* Stellar power spectra (top) and cross correlation coefficients (bottom) for the reference Sherwood-Relics simulation (solid) and the map produced using `JERALD` (dashed) at the same redshifts.

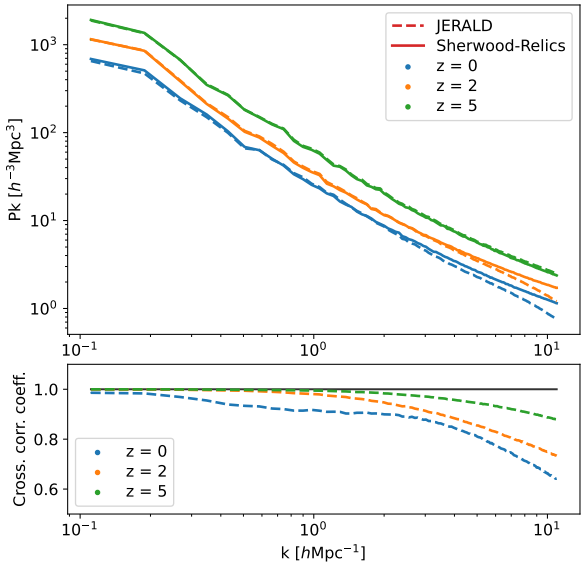


Figure 2.9: Low-intermediate density neutral hydrogen power spectra (top) and cross correlation coefficients (bottom) for the reference Sherwood-Relics simulation (solid) and the map produced using `JERALD` (dashed), at three different redshifts. For visualization purposes, the green curves ($z = 5$) in the top panel were offset by multiplying them by an arbitrary constant (1.7), otherwise they would overlap with the orange ones ($z = 2$). The correct scale of the power spectra at $z = 5$ is the same as for $z = 2$.

2.3.4 Generalization capabilities

Until now, we presented results evaluated on the same reference maps used in the training procedure. Despite the fact that only 50% of the voxels in each map are used for training and the model has only few parameters, its performance in terms of the power spectra and relative differences reported above may still be affected by the parameters specializing on the specific realization on which they are trained; for this reason, in this section we investigate the performance of `JERALD` on unseen simulations. Specifically, we run a new full-hydrodynamical simulation using the same code as the one used for the reference simulation of the previous sections, same box and mesh sizes and same cosmological and astrophysical parameters, but a different seed for the generation of the ICs. We then run FastPM as in Section 2.3.2 with the new ICs and apply `JERALD` with fixed parameters, trained on the simulation used in the previous sections.

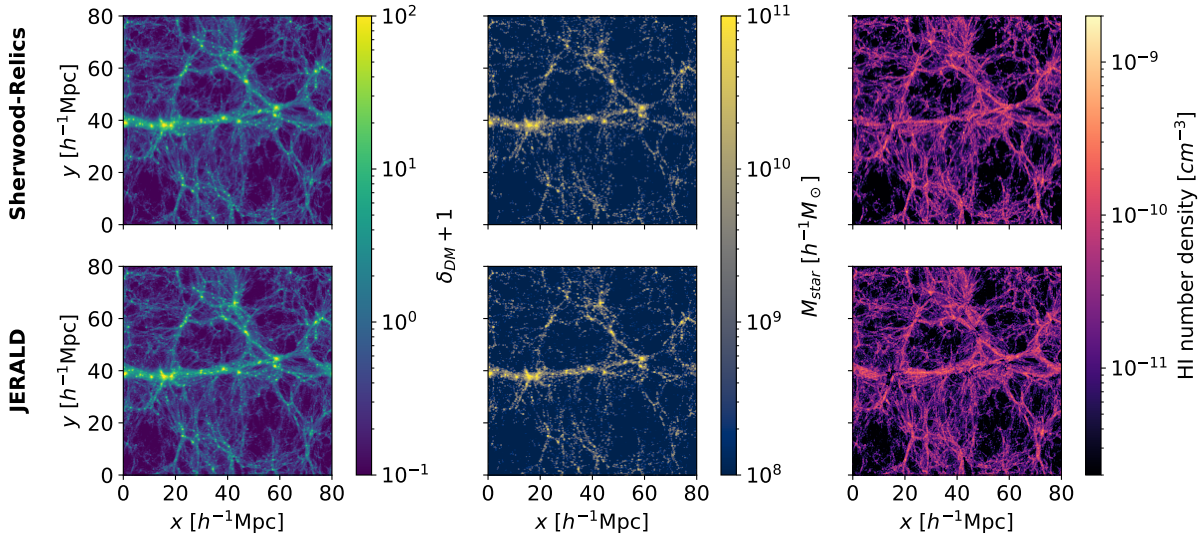


Figure 2.10: Similar to Figure 2.5, but for a new Sherwood-Relics simulation unseen by `JERALD`. In this case, all plots correspond to $z = 0$ and the different quantities—i.e., DM overdensity, stellar mass and HI number density—are arranged horizontally. Here no test sub-box is highlighted, because `JERALD` was pre-trained on a different simulation.

Figure 2.10 shows dark matter overdensity, stellar mass and neutral hydrogen number density maps as in Figure 2.5 with the new unseen simulation. We only report the $z = 0$ maps, as they are the most difficult to model. The visual agreement is good also in this case, with mild discrepancies for HI in the central band. In Figure 2.11 we show the associated power spectra and cross correlation coefficients, including the $z = 2$ case for completeness. We compare the new reference (solid) with `JERALD`'s output (dashed), and we observe similar performance as what was obtained on the reference simulation on which `JERALD` was trained. This is a strong indication of the method's generalization capabilities; this will be expanded upon in the next chapter.

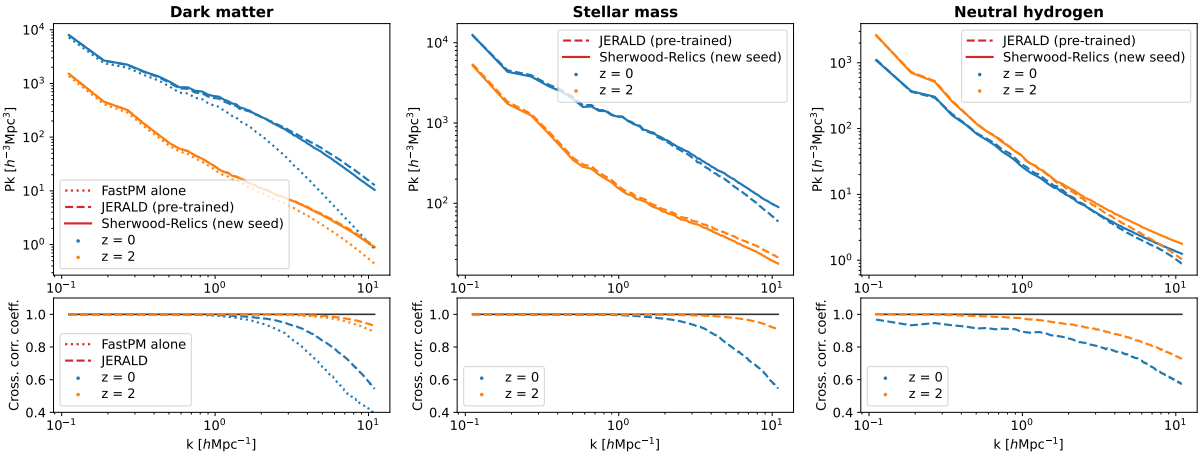


Figure 2.11: Dark matter, stellar and neutral hydrogen power spectra at $z = 0$ and $z = 2$ for a target simulation with box size $(80h^{-1}\text{Mpc})^3$ and mesh size 1024^3 with different initial conditions with respect to the one on which `JERALD`'s parameters were trained, along with the results from `JERALD`.

2.4 Parameters degeneracies

The number of parameters used in this work, specified in Table 2.1, may seem excessive when compared with semi-analytical approaches (e.g. [163]). Despite being a physics-based approach, `JERALD` also needs to correct for effects that do not necessarily have a direct physical interpretation, such as the coarse-graining in space due to the usage of a particle-mesh-based simulator and in time due to the integration scheme employed by FastPM, which requires additional degrees of freedom. Additionally, while semi-analytical models derive expressions for different processes using a specific theoretical model, the goal of `JERALD` is to produce fields using a more general effective approach. Nonetheless, it is possible that some parameters in the current model may be redundant: we observed that different combinations of parameters can lead to very similar losses and performance in terms of summary statistics, suggesting the presence of multiple similar minima (or a global minimum which is rather flat along some degenerate directions), which may be related to the chosen functional forms for the displacements or the final transformations, and it is conceivable that different expressions with fewer parameters could remove possible degeneracies while leading to comparable performance. For example, considering the displacement layers in the case of DM, Eqs. (2.2) and (2.7), we observe that, when using only one layer, different initial parameters can lead to very similar optimized values for γ_{dm} but different ones for α_{dm}, k^h_{dm}, k^l_{dm} and ν_{dm}, although with very similar behaviors of $\tilde{G}_{\text{dm}}(k)$ in the range of frequencies associated to the maps (generally, as one may expect, with greater differences at large frequencies $k \gtrsim 5h\text{Mpc}^{-1}$). As an illustration of this observation, in Figures 2.12 and 2.13 we show the results of two optimizations starting from different initial parameters (reported in Table 2.2) and converging to different parameters

(also reported in Table 2.2) for the DM case discussed in Section 4.3, but with a single displacement layer. Notice that, at convergence, both training and validation losses are within $\sim 1\%$ between the two sets of parameters, and the $\gamma_{\rm dm}$ parameters of Eq. (2.2) are also within $\sim 1\%$. The next chapter will expand on these observations, analyzing them in far deeper detail.

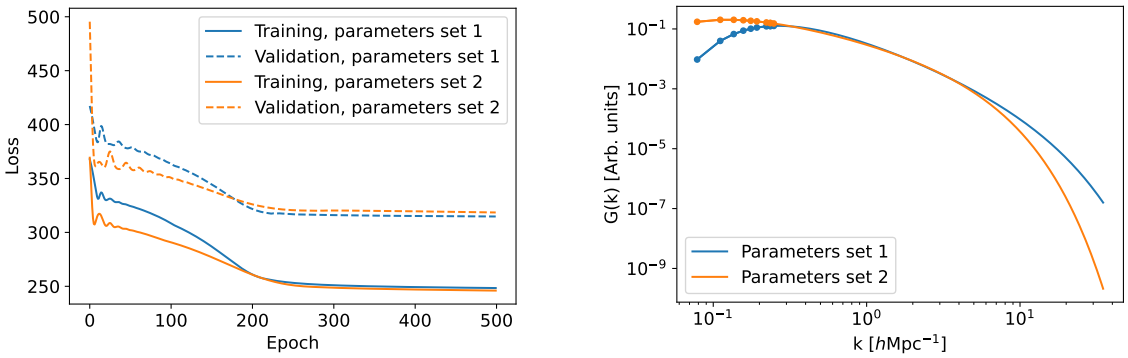


Figure 2.12: *Left:* training and validation losses for two DM optimizations as discussed in Section 4.3, with a single displacement layer, for two different sets of initial parameters. *Right:* Green's function at convergence for the two set of initial parameters, illustrated in the range of frequencies of the associated maps. The dots on the top-left highlight the very few wavenumbers (10) for which the two curves do not overlap at large scales, whereas there are $\gtrsim 3 \times 10^5$ values of k above $\sim 5h$Mpc^{-1} for which the curves significantly differ.

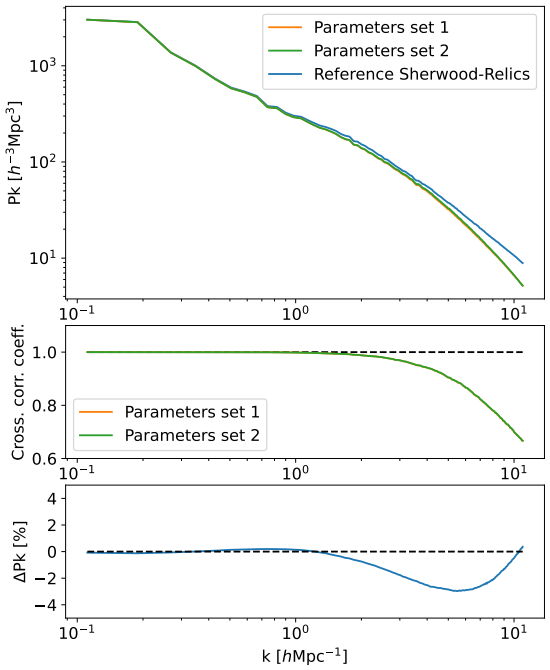


Figure 2.13: Final power spectra (top), cross correlation coefficients (middle) and power spectra relative difference (bottom) for two DM optimizations as discussed in Section 4.3, but with a single displacement layer, for two different sets of initial parameters. The curves overlap within a few percent.

Table 2.2: Initial parameters and their values at convergence for two different optimizations in the case of DM discussed in Section 4.3, but with a single displacement layer.

	Initial values		At convergence	
	Set 1	Set 2	Set 1	Set 2
$\alpha_{\rm dm}$	0.001	0.002	0.06775	0.053
$\gamma_{\rm dm}$	0.5	0.8	0.61536	0.62374
$k^h_{\rm dm}$	1	0.5	0.58686	0.16695
$k^l_{\rm dm}$	10	5	6.64337	2.37339
$\nu_{\rm dm}$	0	0	-2.17048	-1.31428

2.5 Conclusions

We presented a new, improved implementation of a physically motivated approach to reducing the enormous computational cost of cosmological simulations in order to obtain accurate dark matter and baryonic maps. Our code, `JERALD` (JAX Enhanced Resolution Approximate Lagrangian Dynamics), builds on the Lagrangian Deep Learning method originally proposed by [149], and focuses in particular on producing, on top of more accurate DM and stellar maps, high-fidelity neutral hydrogen maps.

We showed that our improved implementation is capable of producing maps in excellent agreement, in terms of power spectra for DM, stars and HI, with full-hydrodynamical simulations, up to several $h\text{Mpc}^{-1}$. The performances of `JERALD` are more impressive at higher redshift ($z \geq 2$), where non-linearities are less prominent: in this regime, the agreement, in terms of power spectra, is above the 70% level all the way to very small scales ($k \gtrsim 10\,h\text{Mpc}^{-1}$). This opens the door to investigating, in quantitative detail, the impact on very small scales of different dark matter models.

`JERALD` was developed from scratch, with the aim of being user-friendly, fully parallel and highly computationally efficient. Its ultimate goal is not limited to producing baryonic maps: we aim at exploiting this approach to study observable HI-related quantities such as the Lyα forest and the 21cm intensity [154]. These two observables are particularly important, as neutral hydrogen is an excellent probe of DM models beyond cold dark matter both in emission and absorption [164, 159]—an alternative that could in principle alleviate some of the small-scale inconsistencies of the ΛCDM model: for example, a lighter ("warm") DM particle could offer the solution to the core/cusp problem, as the main effect of its larger velocity is to "smear out" structures at small scales [165, 166]. The quantitative study of such models from large scale structure and Lyα forest, as well as 21cm mapping, can benefit from being embedded in a neural network-enabled simulation-based inference (SBI) framework (e.g., [167, 168, 169, 170]), full high-dimensional likelihood evaluation [171, 172] or field-level inference (e.g. [173]) approaches, which have been shown to extract a great deal more information from the data than traditional hand-crafted sum-

maries such as the power spectrum. Often, the bottleneck to the full exploitation of such techniques lies in the limited amount of high-resolution simulations on which to train the neural models. With `JERALD`, we have introduced a tool that will enable us to take a major step in future work to enable such studies.

Chapter 3

JERALD as an emulator

All results presented in the previous chapter were obtained for fixed cosmology and astrophysics. However, the end goal of JERALD is to serve as part of an emulator, e.g. within an SBI pipeline, to eventually perform fast field-level inference on cosmological and astrophysical parameters. Ideally, the emulator would consist of three components: a Boltzmann solver, e.g. CLASS [41], which generates the linear power spectrum; an approximate N-body simulator, e.g. FastPM, which takes Ω_m, σ_8, and the linear power spectrum as inputs to simulate an approximate DM distribution at the required redshift; and finally JERALD, to correct the DM distribution produced by the approximate simulator and, optionally, paint baryons on top. A schematic of this pipeline is shown in Figure 3.1.

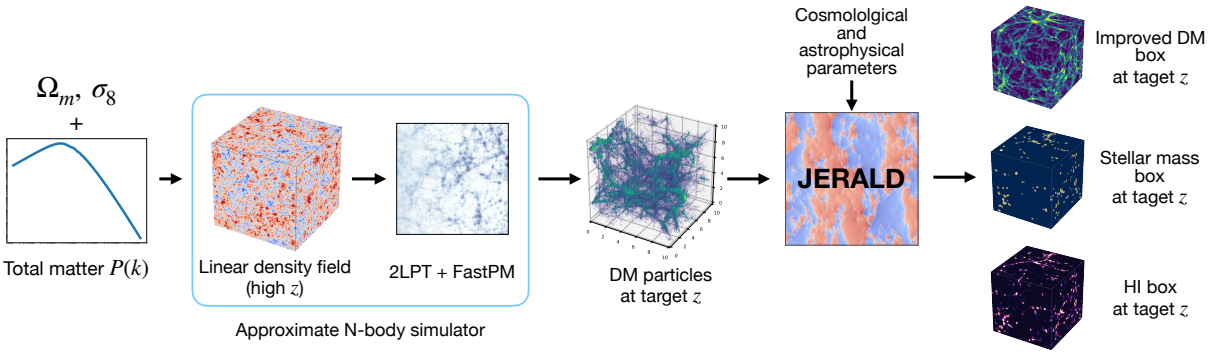


Figure 3.1: Pipeline of JERALD used as an emulator, together with the linear matter power spectrum generation and the approximate N-body simulator (FastPM).

To achieve this, JERALD must be able to produce results that depend on cosmological and astrophysical parameters. This chapter is therefore dedicated to assessing whether this goal is achievable in practice — that is, without training a separate set of model parameters $\boldsymbol{\theta}$ for each choice of cosmological/astrophysical parameters $\boldsymbol{\phi}$. Two strategies could be adopted for this purpose.

The first is to condition JERALD's architecture on $\boldsymbol{\phi}$ directly, so that the model takes

the cosmological and astrophysical parameters as inputs; this is pursued in Section 3.1, where `JERALD` is modified to either generate $\boldsymbol{\theta}$ as the output of a neural network or by replacing the Green's function with a spline to achieve the same goal.

The second is to leave the model untouched and ask whether there exists a functional relation $\boldsymbol{\theta}(\boldsymbol{\phi})$ that could be characterized once and then simply evaluated, rather than re-fitted. If such a relation exists and is smooth enough to be interpolated, it would deliver the same practical outcome as a conditioned `JERALD`, i.e. a model applicable to a new cosmology without a new optimization, while additionally preserving the physical interpretability of the original parametrization, and possibly offering additional insight into the physical meaning of the parameters themselves. The rest of the chapter is dedicated to addressing this question. In practice, however, the existence of such a relation presupposes that $\boldsymbol{\theta}$ is well determined at fixed $\boldsymbol{\phi}$: i.e. that any optimization returns, up to small numerical scatter, a unique answer for a given target simulation. As already suspected in the work presented in the previous chapter (Section 2.4) and the sections that follow establish directly, this presupposition does not hold: the search for a relation $\boldsymbol{\theta}(\boldsymbol{\phi})$ is then intertwined with the study of the model's own loss landscape, with the behavior of its parameters and with what the optimization is actually determining. Specifically, Section 3.2 varies redshift alone, while keeping cosmology, astrophysics and initial conditions fixed, and finds the recovered parameters already poorly behaved in this comparatively benign setting. Section 3.3 turns to the case of interest, in which $\boldsymbol{\phi}$ varies, first attempting a direct regression of $\boldsymbol{\theta}$ on $\boldsymbol{\phi}$, although recovering nothing better than the sample mean. The reason is isolated by examining a dataset in which cosmology is held fixed and only the seed of the initial conditions changes: the parameters scatter both between realizations and, at fixed realization, between independent optimizations, the latter tracing out the degeneracy directions anticipated in Sections 2.4 and 3.2. Following those degeneracies to their origin identifies a combination of the parameters and of the input field, $J(k)$, as the quantity the loss actually constrains, together with a residual sensitivity on the exponent of the source term γ. With these well-defined targets, Sections 3.4 and 3.5 test whether they can be predicted. The first compares γ against statistics measured on the input field, motivated by a physical argument for what it should be responding to, recovering a real but weak dependence on the realized clustering amplitude. The second attempts direct regressions on both γ and $J(k)$, finding ceilings for the amount of variance that can be predicted, in each case set by the information content of the inputs, and provides explanations for why those ceilings arise. Since a usable mapping $\boldsymbol{\phi} \mapsto \boldsymbol{\theta}$ cannot be identified, Section 3.6 asks the complementary question of how much is lost by not having one, and applies a single fixed parameter set to simulations it was never fitted to; the answer proves favorable at the level of two-point statistics and of the mean, while less so for individual realizations. Finally, Section 3.7 identifies what limits the performance attainable on a

given simulation at all, and Section 3.8 collects the outcome.

The next sections will focus exclusively on the DM case; as such, the subscripts "dm" used in the definitions of the model parameters in the previous chapter will be dropped. Note that $k^l_{\rm dm}$ and $k^h_{\rm dm}$ will be referred to as k_l and k_h respectively (with subscripts rather than superscripts). Additionally, as mentioned in Section 2.3.2, α enters the Green's function multiplied by a factor 10^{-2} in the code, therefore the values reported in this chapter should be read accordingly.

3.1 Conditioned JERALD

3.1.1 Neural JERALD

A simple idea to allow JERALD to take $\boldsymbol{\phi}$ as inputs consists in integrating a neural network (NN) into the model, defining the model parameters $\boldsymbol{\theta}$ as its outputs $\boldsymbol{\theta} = \mathrm{NN}(\boldsymbol{\phi})$ and training the NN weights rather than $\boldsymbol{\theta}$ directly. One challenge associated with this approach is that training such a model would require using a large number of target maps (for different $\boldsymbol{\phi}$) and running the model for many of them at once, making optimizations both time and memory extremely intensive. Of course, this issue could be mitigated by drastically decreasing the resolution of target maps, although at the cost of losing detail at small scales. Using target simulations with small box size could allow to reduce the map sizes considerably without discarding small-scale structures; however, small simulation boxes amplify the effect of cosmic variance, which, as discussed in Sections 3.3.3 and beyond, considerably complicates establishing a clean dependence of JERALD's parameters on $\boldsymbol{\phi}$, undermining the very property this chapter aims at achieving. The usage of a NN version of JERALD *for a single target map* (the same as in the previous chapter) was briefly investigated, although in every case resulting in training divergences and instabilities due to the excessive freedom of the network's outputs, and the approach was eventually discarded. This outcome is expected: the parameters $\boldsymbol{\theta}$, being physical, cannot be arbitrarily large or small, and some of them, namely k_h and k_l (see e.g. Eq. (2.7)), cannot change sign; a NN with unconstrained outputs then naturally produces divergences. Imposing the proper constraints to avoid this phenomenon and initializing the NN weights such that the initial guesses for $\boldsymbol{\theta}$ are sensible, however, is not straightforward. Naturally, one could simply apply special transformations to the NN outputs (e.g., sigmoid to rest rict their ranges and/or softplus to enforce positivity) and, for the most part, ignore specific weight-initialization strategies; however, on top of potentially affecting the NN training, this is a largely arbitrary operation, likely requiring precise tuning of many hyperparameters and possibly limiting the quality of the results. This alternative was not tested.

3.1.2 Spline Green's function

A different approach to integrate a NN into the model and allow conditioning on parameters, while at the same time potentially make `JERALD` more flexible, loosely inspired by [91], consists in defining a new Green's function as a B-spline [174], whose parameters are the outputs of a small and simple feed-forward neural network taking $\boldsymbol{\phi}$ as inputs. Concretely, B-splines are smooth piecewise polynomials of arbitrary degree, defined by a set of knots (values on the x axis) and control points (values on the y axis). The implementation follows [91], drawing in particular on the associated repository. As a first, simplified attempt, γ of Eq. (2.2) was fixed to unity and $\tilde{G}(k)$ was defined directly as a B-spline rather than as in Eq. (2.7), i.e.

$$\tilde{G}(k) = \text{B-spline}(k,\, t,\, c,\, d)\ , \tag{3.1}$$

where t and c, the outputs of the neural network, are the knots and control points respectively, and d is the polynomial degree (set to 3, as in [91]). The number of knots and control points was initially set arbitrarily to 8. At first, the neural network input was set to the scale factor: since training was performed at fixed redshift $z = 0$, this amounted to a constant input of 1. To simplify the problem further and isolate the flexibility of the B-spline itself, the network was subsequently removed entirely, and the knots and control points were instead treated directly as the effective parameters to be learned, following the same procedure used previously for `JERALD`. Some example results are shown below.

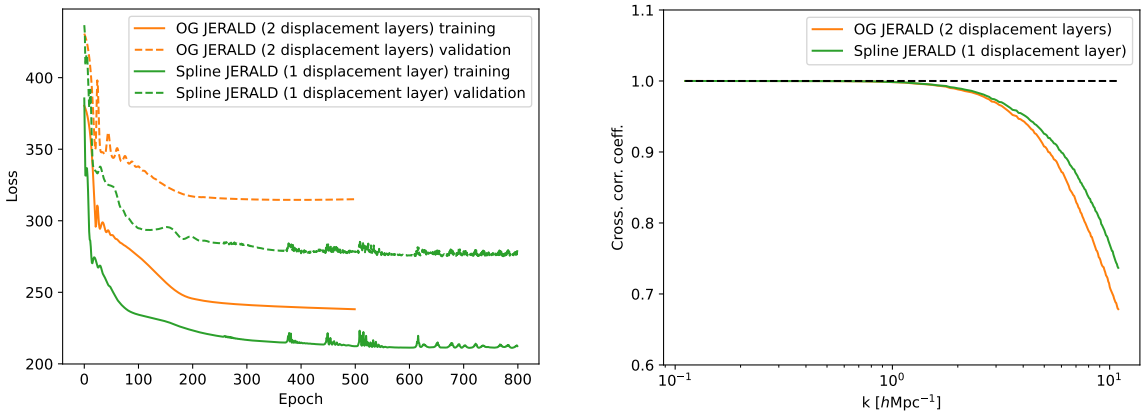


Figure 3.2: Training and validation losses (*left*) and cross correlation coefficients (*right*) for the original `JERALD` (OG `JERALD`) with 2 displacement layers and the version with the B-spline with a single displacement layer, both trained in the same setting and on the same Sherwood-Relics target as in Chapter 2.

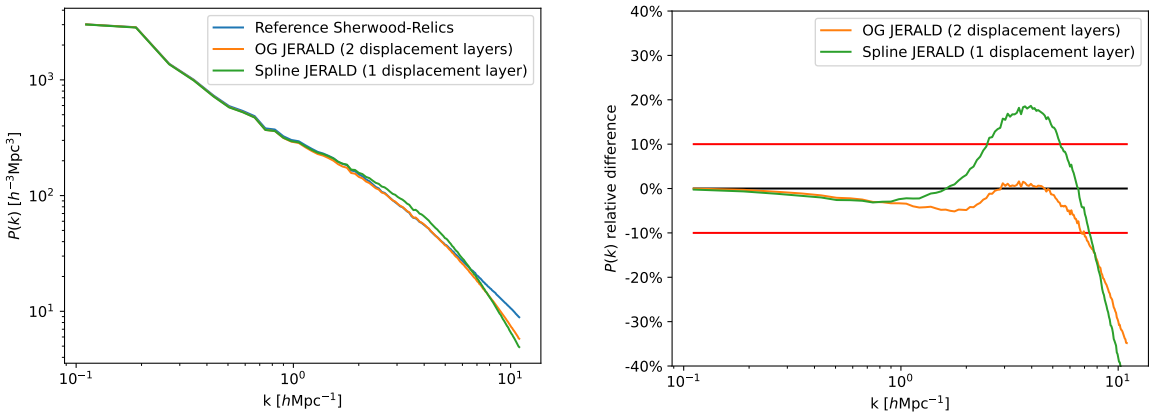


Figure 3.3: Power spectra (*left*) and their relative differences with respect to the reference (*right*) for the original `JERALD` (OG `JERALD`) with 2 displacement layers and the version with a single B-spline displacement layer, both trained on the same Sherwood-Relics target.

These optimizations generally lead to better losses with respect to the original `JERALD` (for example going from final training losses of ~ 240 to below 200 for the same target map), even with a single displacement layer. This also translates into better cross correlations, as Figure 3.2 shows; however, the power spectrum is worsened with respect to the original `JERALD`, as shown in Figure 3.3. This may indicate that the choice of the Green's function in Eq. (2.7) is particularly well-suited to this type of problem, which could be reasonable considering the interpretation of its functional form, provided in the previous chapter.

Many other options (and combinations of them) with the B-spline Green's function, briefly listed here, were tested:

- 2 displacement layers, with the Green's functions either sharing or having different parameters among the layers;
- Reintroducing the exponent γ in Eq. (2.2) as an additional parameter to learn together with the knots and control points;
- B-spline Green's function divided by different powers of the wavenumber, to also allow negative powers of k;
- Green's function defined as the exponential of a B-spline with a negative sign, divided by different powers of the wavenumber, in order to align it with the exponential part of Eq. (2.7);
- Original Green's function with an additive term given by a B-spline, optionally with a tunable coefficient, in order to retain the statistical properties of the original Green's function while at the same time allow for increased flexibility;

- Combinations of a B-spline Green's function and one or more terms from the original Green's function with k_l, k_h or ν;
- Different numbers of knots and control points;
- Fixed knots, logarithmically spaced in the wavenumbers range of the density fields;
- Different boundaries for the control points, forcing $G(k) \to 0$ with different powers of k for $k \to 0$ and $k \to k_{max}$, where k_{max} is the largest wavenumber associated to the density fields;
- Different losses with respect to the original L2, such as relative losses or L1 loss.

In general, as in the example shown above, these changes lead to improved losses and cross correlations with respect to the original `JERALD`, but in every case with worse statistics as given by the power spectrum. To further check the quality of the statistics of the output fields, the bispectra (see Eq. (1.33)) of the maps produced by the original `JERALD` and the version with the plain B-spline Green's function were also computed, in this case highlighting improvements in some cases and worse performance in others; some example plots are included in Figures 3.4, 3.5 and 3.6.

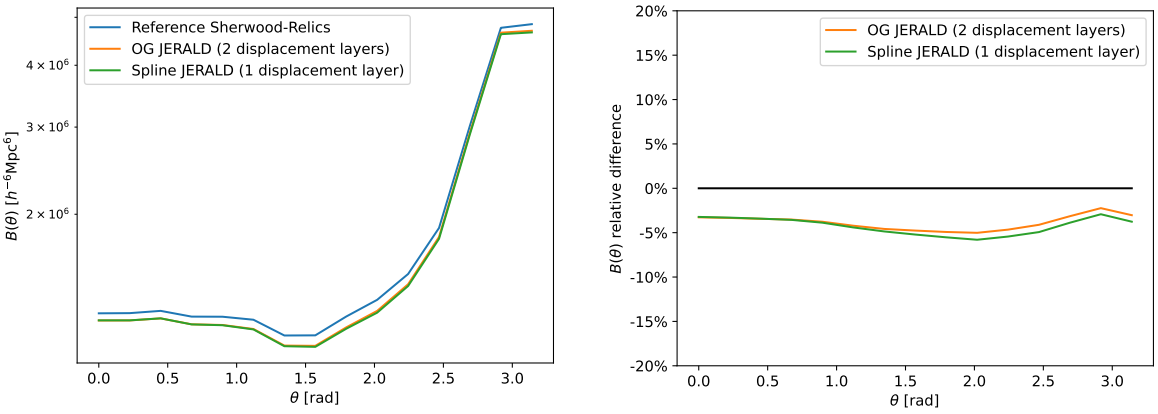


Figure 3.4: Bispectra (*left*) and their relative differences (*right*) with respect to the reference for $k_1 = k_2 = 0.5\,h\mathrm{Mpc}^{-1}$ for the original `JERALD` (OG `JERALD`) with 2 displacement layers and the version with the B-spline in the same configuration as for Figures 3.2 and 3.3, both trained on the same Sherwood-Relics target.

Eventually, the addition of a power spectrum term in the loss itself was tested (see Appendix C for the implementation of the power spectrum and its derivative), yet still leading to no improvement. However, no systematic exploration of the hyperparameter weighting this term against the voxel-space loss was performed: therefore, this remains an open avenue for further investigation. Naturally, in this case other summary statistics

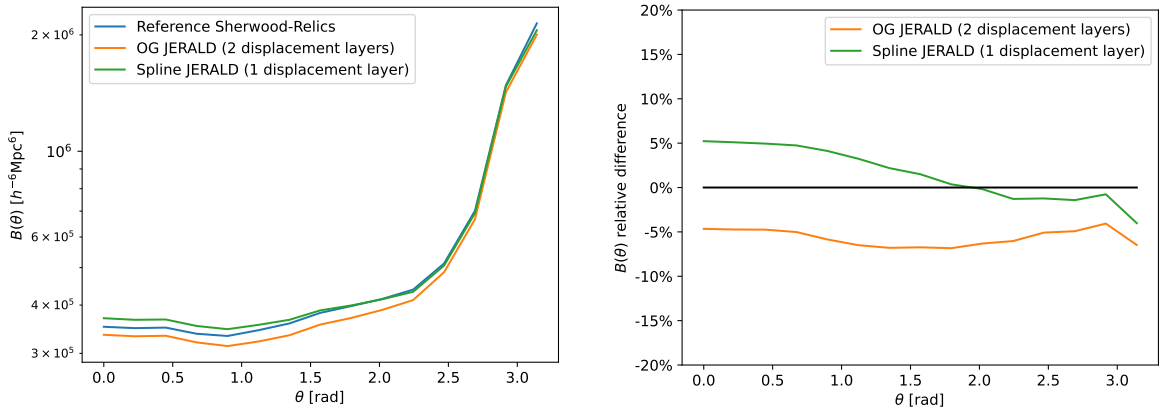


Figure 3.5: Same as Figure 3.4, but for $k_1 = k_2 = 1\,h\mathrm{Mpc}^{-1}$.

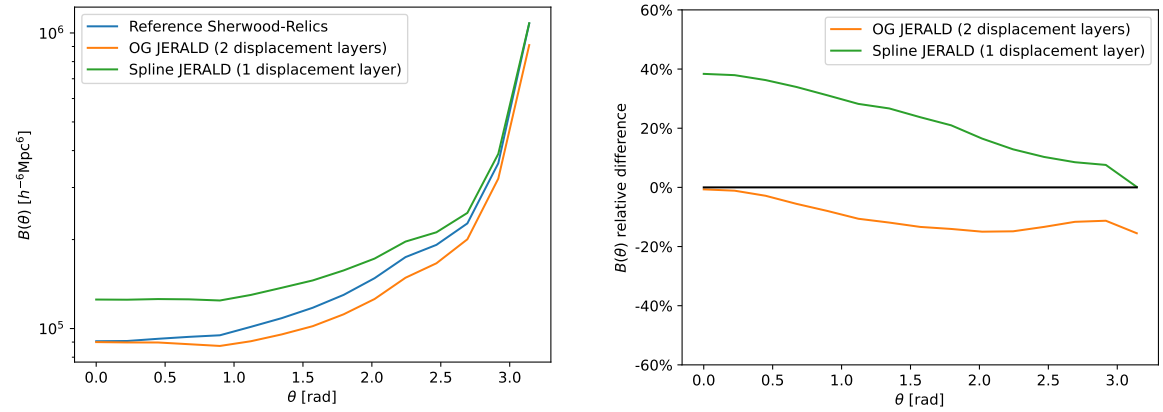


Figure 3.6: Same as Figure 3.4, but for $k_1 = k_2 = 2\,h\mathrm{Mpc}^{-1}$.

with respect to the ones used in the work described in the previous chapter, such as the bispectrum, would need to be evaluated in order to assess the quality of the output fields.

Overall, the pattern across every variant tried is consistent: as mentioned above, added flexibility in the Green's function reliably improves the voxel-space loss and cross-correlation coefficient, but at the same time worsening the power spectrum, while the bispectrum comparisons (Figures 3.4 through 3.6) show no clear trend. A possible reading is that Eq. (2.7) is not simply one convenient parametrization among many: its asymptotic behavior is tied to the gravitational-potential limit recovered at $\alpha = 4\pi G\bar{\rho}$, $k^h = 0$, $k^l = \infty$ and $\nu = -2$ (see Eq. (1.35)), and a generic spline has no built-in reason to preserve such behavior even when it fits the training loss better. Minimizing an L2 loss in voxel space and reproducing the correct clustering statistics are not necessarily the same objective. This does not rule out the spline approach, but it does suggest that following it would introduce the additional requirement of anchoring the function with the same physical motivations that make Eq. (2.7) work, rather than letting the knots and control points

vary freely. The boundary-constrained variant listed above and the one retaining the original Green's function with an additive spline term, combined with a properly tuned power spectrum loss term, remain thus more promising routes than treating the spline as an unconstrained fit.

Note that the setup in [91] differs from that of `JERALD` in several aspects: there, the reference is a full N-body simulation rather than a full-hydrodynamical one, so that the residual to be modeled is purely gravitational; additionally, the filter enters the gravity solver itself, alongside the physical model and Ω_m, and acts at every integration step rather than only at the target redshift, with an explicit dependence on the scale factor; finally, the loss optimizes the distance between the output and reference particles positions rather than fields, together with the relative power spectrum. Despite these differences, some insights from [91] may still be beneficial in case this direction for `JERALD` is pursued: [91] states that the power spectrum term in the loss was introduced precisely because penalizing positions alone failed to recover the small-scale power spectrum, which is the failure mode encountered above; moreover, their filter only parametrizes a departure from the physical solution rather than replacing it outright. Both features are the remedies suggested above. Nonetheless, the connection between the two approaches is very loose, and therefore a remedy effective in one may not be effective in the other; moreover, the time and memory limitations discussed in Section 3.1.1 still pose a significant challenge. For these reasons, it is worth exploring an alternative path to make `JERALD` capable of generalizing for different inputs.

3.2 Redshift dependence

Following the unsuccessful attempts of the previous sections at developing a version of `JERALD` that allows conditioning on inputs, the next sections leave `JERALD` unchanged, focusing on finding a mapping between its optimized parameters $\boldsymbol{\theta}$ and the cosmological and astrophysical parameters $\boldsymbol{\phi}$ it should depend on to be a usable emulator for a statistical inference pipeline. Before considering changes in $\boldsymbol{\phi}$ and possibly also initial conditions, however, it is useful to vary a single, continuous quantity: loosely inspired by Section 4 of [90], a test was carried out to establish whether there exist a redshift dependence of $\boldsymbol{\theta}$, $\boldsymbol{\theta}(z)$, in the simplest case, i.e. a single displacement layer for DM (5 model parameters). For this purpose, `JERALD` was tested in the same setting as Chapter 2 on 25 snapshot of a Sherwood simulation [77] (with constant $\boldsymbol{\phi}$), from redshift 8 down to 0, approximately linearly spaced in scale factor. The model's high sensitivity to initial conditions observed in the work described in the previous chapter leads to the speculation that the loss landscape may feature multiple local minima and/or very flat regions: for this reason, at first each optimization (for each redshift) was performed starting from the same initial guesses

for $\boldsymbol{\theta}$. However, this lead very early to erratic behaviors of the optimized parameters, as no relation $\boldsymbol{\theta}(z)$ could be identified. Therefore, following the assumption that this relation does exists, new optimization were run, each using as initial guesses for $\boldsymbol{\theta}$ the optimized ones at the "previous" redshift (e.g. optimizing the parameters at $z = 6$ starting from the optimized $\boldsymbol{\theta}(z = 7)$). The direction of this redshift sweep was chosen to be from $z = 8$ downwards, as at higher redshift the magnitude of the correction that `JERALD` has to apply should be small, potentially making it easier to notice critical points in which the behavior of $\boldsymbol{\theta}(z)$ changes significantly. The results of this set of optimizations are showed in Figure 3.7: visually, the curves $\boldsymbol{\theta}_i(z)$ appear relatively smooth down to the leftmost point at $z = 3.2$, where an abrupt fluctuation of the parameters occurs; this behavior is also manifested in the Green's functions, as displayed on the right panel of Figure 3.7, where the lime curve ($z = 3.2$) exhibits a significantly different trend compared to the other curves. Note that both plots stop at $z = 3.2$ as the redshift sweep was not continued beyond that point; Section 3.2.2 analyzes a redshift range around $z = 3.2$ more in detail.

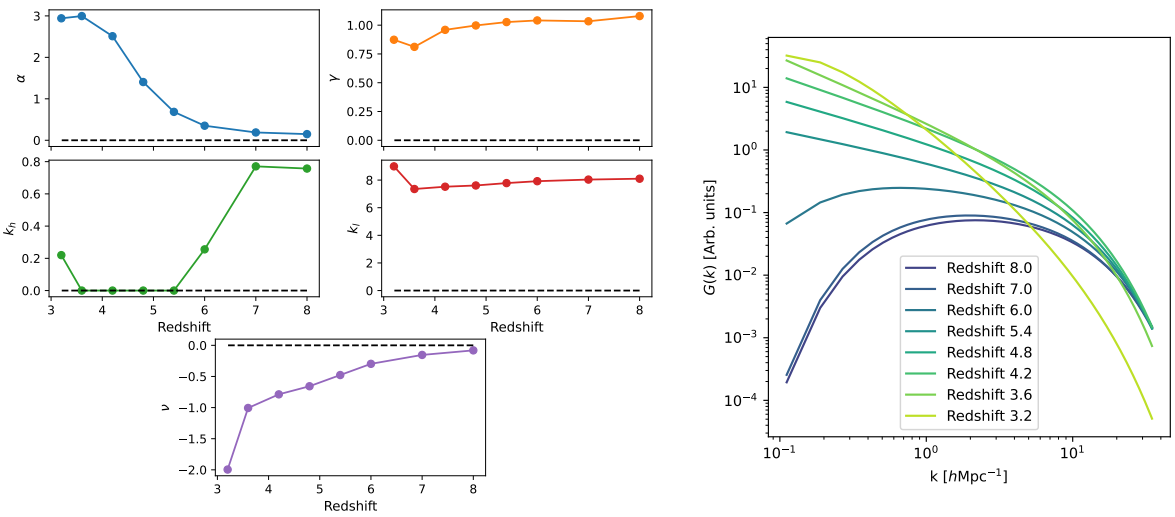


Figure 3.7: *Left:* behavior of the model parameters as functions of redshift, with optimizations starting from the optimized parameters at a previous redshift (e.g. the initial guesses for the parameters at $z = 6$ are the optimized parameters for $z = 7$). *Right:* corresponding Green's functions.

The same procedure was also attempted in the opposite direction, sweeping from $z = 0$ upwards. In this case, an anomaly appears already at the second redshift considered, $z = 0.1$: over the first optimization steps the validation loss oscillates strongly, and it subsequently stabilizes at, and then slowly drifts upwards from, a value appreciably higher than the minimum it had transiently reached during the oscillatory phase. In other words, the optimization does visit better regions of parameter space than the one in which it eventually settles, but does not remain there. This behavior proved insensi-

tive to the learning rate and its schedule, to the remaining optimizer hyperparameters, and to the choice of optimizer itself, being reproduced with Adam [161], AdamW [175], RMSprop [176], plain gradient descent and gradient descent with momentum. Given the small number of parameters, second-order and non-gradient alternatives were also tested, namely the BFGS algorithm [177] and a least-squares solver, both of which performed worse than Adam. The fact that the effect survives every change of optimizer suggests that it originates in the loss landscape rather than in the optimization scheme: this is consistent with the multiple minima and flat directions documented later in this section, and with the possibility that the minimum found at $z = 0$ is a poor starting point from which to track the parameters upwards in redshift.

3.2.1 Redshift evolution of the input and reference simulations

To understand the origin of the behavior at $z = 3.2$, a first step is to establish whether it can be traced to the simulations themselves, rather than to the model. Two mechanisms could plausibly produce a discontinuity in the optimized parameters at a particular redshift: an abrupt change in the DM dynamics between consecutive snapshots, caused, for instance, by the formation of an unusually massive halo displacing large numbers of surrounding particles, or a change in the way baryonic feedback in the reference simulation imprints itself on the DM distribution, which the N-body simulator, lacking baryons entirely, cannot reproduce.

The first mechanism can be tested directly. A large and coherent redistribution of particles between two snapshots would reduce their statistical similarity, so the cross-correlation coefficient between consecutive Sherwood DM maps provides a convenient proxy. The left panel of Figure 3.8 shows this coefficient for the pairs $z = 4.2 - 3.6$ and $z = 3.6 - 3.2$: the two curves are closely comparable, with the latter pair slightly more correlated than the former, as expected given the smaller redshift interval separating them. There is thus no indication of anomalous structural evolution across the redshift at which the parameters change behavior.

A more direct probe of the same mechanism is the mean displacement of particles between consecutive snapshots, measured for both the reference and the approximate simulation and illustrated in Figure 3.9. Unlike the cross-correlation, which samples only two snapshot pairs, this diagnostic covers the entire range from $z = 8$ to $z = 0$, and can therefore establish whether any redshift in the sweep is anomalous.

As the figure shows, the mean displacement grows steadily as redshift decreases, as expected from the progressive growth of structure; the apparent sharp drops in the left panel are artifacts of the varying size of the redshift steps between stored snapshots, and the right panel of Figure 3.9 shows the same quantity after dividing by the step size in order to expose the underlying trend (the left panel was kept as it illustrates an actual

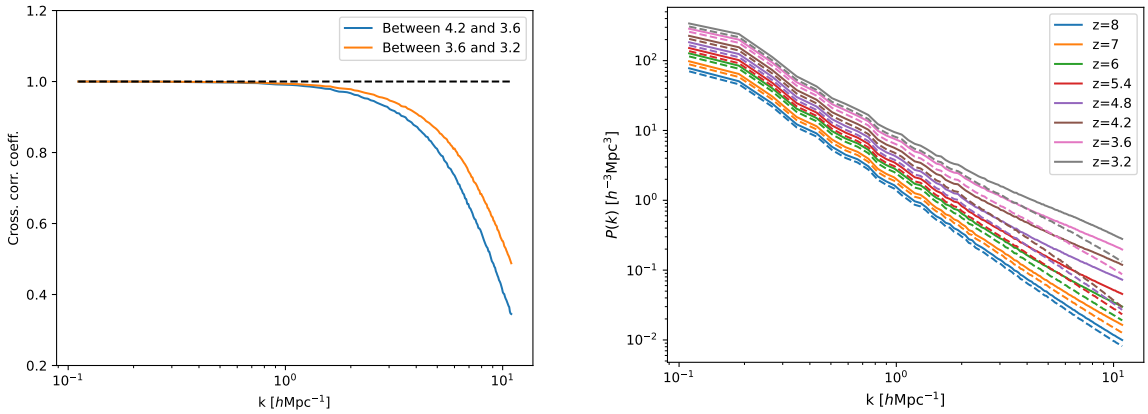


Figure 3.8: *Left:* cross-correlation coefficients between Sherwood snapshots at $z = 4.2$ and 3.6 and between $z = 3.6$ and 3.2. *Right:* DM power spectra for $z = 8$ down to 3.2 for Sherwood (solid lines) and the approximate N-body simulator (FastPM, dashed lines).

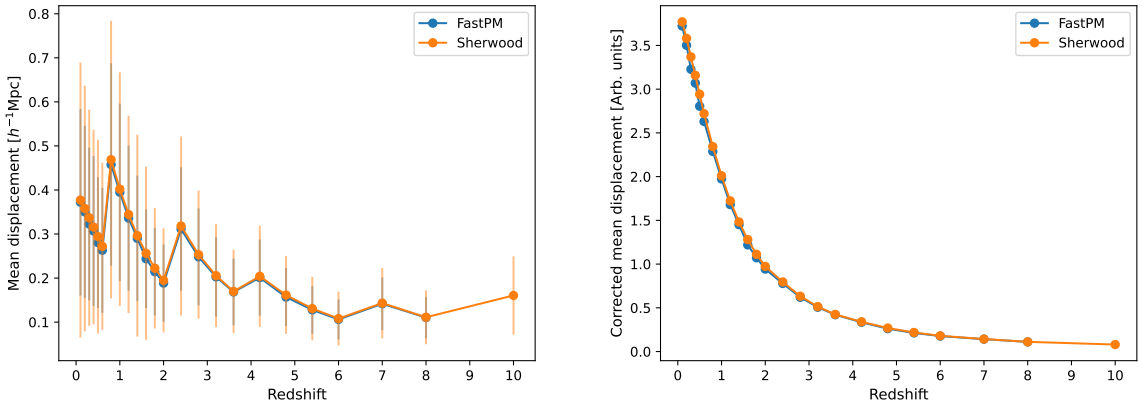


Figure 3.9: Mean displacement with bars indicating $\pm 1\sigma$, unprocessed (*left*) and corrected to account for the size of the redshift steps (*right*) to show the underlying displacement pattern. Each point represents the average displacement between the indicated redshift and the one to its left (e.g. the point at $z = 10$ represents the mean displacement between snapshots at $z = 10$ and $z = 8$).

displacement). In neither panel, and in neither simulation, does $z = 3.2$ stand out from the surrounding redshifts. Finally, since the Green's functions of Figure 3.7 are defined in Fourier space, it is also worth comparing the DM power spectra of the reference and approximate simulations across the same range, shown in the right panel of Figure 3.8; here too the progression with redshift is smooth, with no feature singling out $z = 3.2$.

The second mechanism is harder to exclude. In Sherwood, the adopted models for active galactic nuclei (AGN) feedback and for the ionizing background may contribute to this behavior: in fact, these are known to produce a prominent feature in the thermal history of the intergalactic medium (IGM) at around $z = 3$ (see Figure 1 of [77]), and while the impact of such choices is usually assessed on the baryons, they are also known to affect

the total matter distribution appreciably [178, 179, 180, 181]. Whether an effect of this kind is large enough to displace the optimum of a five-parameter effective model discontinuously is, however, not something the present analysis can determine: it would require comparing optimizations against reference simulations with different feedback prescriptions, which was not attempted here. Nonetheless, the model-side explanations examined in the remainder of this section are on their own sufficient to account for the observed discontinuity, so that a baryonic origin may not need to be invoked in order to explain it.

3.2.2 Optimization properties and loss landscape

Assuming the simulations are the less likely origin of the discontinuous behavior in Figure 3.7, attention then turned to the model itself. A natural candidate explanation is the presence of multiple local minima in the loss, already hypothesized in Chapter 2: an optimization may terminate in a sub-optimal minimum, and the global optimum may itself migrate between minima as redshift changes, producing an abrupt shift in the recovered parameters between adjacent snapshots. To test this, 10 independent optimizations were run at each redshift $z = 5.4$, 4.8, 4.2, 3.6, 3.2, 2.8, 2.4 and 2, in the same configuration used above, i.e. a single displacement layer for DM with five model parameters, each starting from a different set of initial guesses drawn randomly within physically sensible ranges.

Figures 3.10–3.15 collect some of the redshifts tested, confirming the presence of multiple minima and showing that their structure evolves with z. All plots feature at least one minimum at $\alpha \sim 1$ (Figure 3.11), with additional optima at varying α (~ 50 at $z = 3.6$, ~ 7.5 at $z = 3.2$ and ~ 15 at $z = 2.4$); contrarily to the other redshifts, $z = 3.2$ appears to prefer the higher-α optimum in terms of validation loss (Figure 3.10), and optimizations seem to be able to escape from the $\alpha \sim 0$ one, although settling around a third minimum (around $\alpha \sim 10$), visible in figure Figure 3.10 and 3.11. Nonetheless, for $z = 2.4$, whether the lower minimum is also preferable in terms of the output statistics is far less clear: as Figure 3.15 shows, the two minima yield power spectra that are accurate over complementary ranges of scale, one agreeing better at intermediate wavenumbers ($\sim 1\,h\mathrm{Mpc}^{-1}$) and worse at smaller scales, and vice versa for the other. This is a further instance of the phenomenon already encountered in Section 3.1.2, where improvements in the voxel-space loss did not translate into improved clustering statistics. The fact that the two minima are complementary rather than simply unequal also suggests that neither is capturing the full range of scales the model is asked to correct, and that a functional form with additional freedom at intermediate wavenumbers (or, more speculatively, some combination of the two solutions) might outperform both; this was not pursued.

The same set of optimizations also exposes a degeneracy between k_l and ν. Rather than clustering around isolated points, the recovered values trace out one or more one-

dimensional curves in the (k_l, ν) plane (Figure 3.13), along which the loss varies only marginally, indicating near-flat directions in the landscape. The effect is equally visible in function space: markedly different combinations of k_l and ν produce almost indistinguishable Green's functions over the relevant range of wavenumbers (Figure 3.14, particularly at $z = 4.8$). The origin of this degeneracy can be attributed to Eq. (2.7), in which k_l and ν both act to suppress $G(k)$, the former through an exponential cutoff and the latter through a power law, so that, over the limited range of scales resolved by the target simulation, the two suppressions become largely interchangeable; this will be returned to in Section 3.3.3. Although the degeneracy curves of Figure 3.13 can be fitted individually, the resulting relations are themselves redshift-dependent, so the degeneracy cannot be removed, nor one of the two parameters eliminated, by a single reparametrization. Finally, the location of the optimum along these curves shifts substantially from $z = 3.6$ onward, most visibly in k_l, and the corresponding Green's functions differ markedly between the two competing minima (see $z = 2.4$ in Figure 3.14), reinforcing the picture of an optimum that relocates with redshift rather than tracking a smooth trajectory.

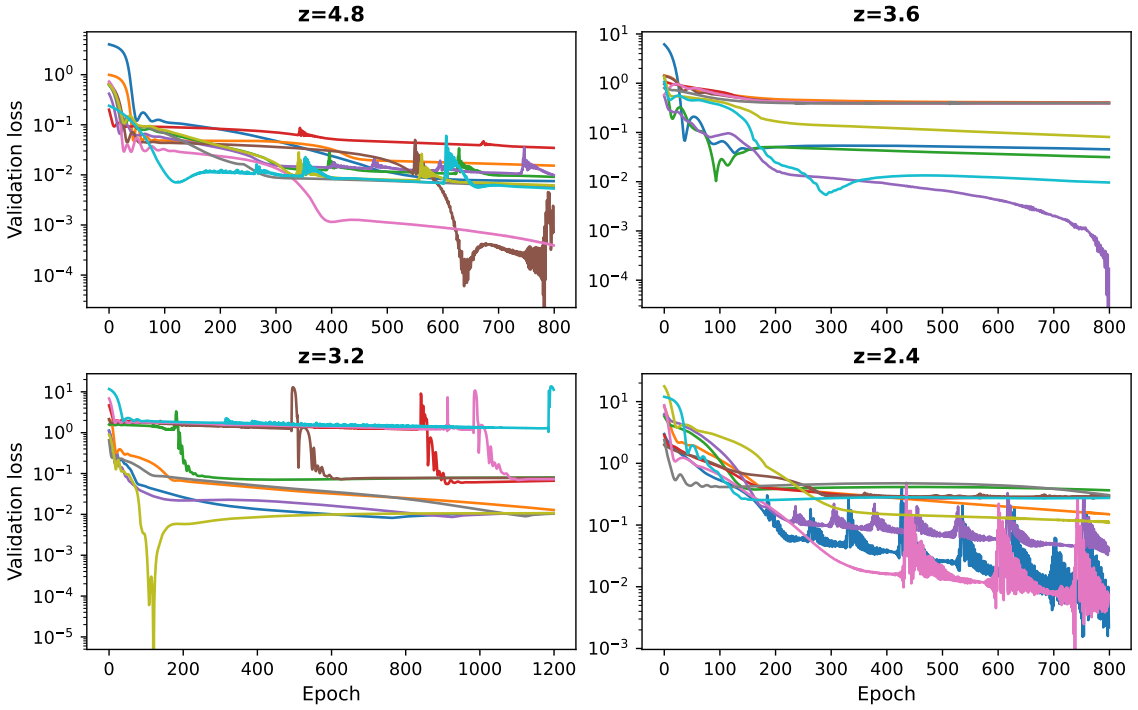


Figure 3.10: Validation losses (with the overall minimum subtracted) through 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4 as functions of training epoch. More optimizations epochs were used for $z = 3.2$ in order to better observe the switching between minima.

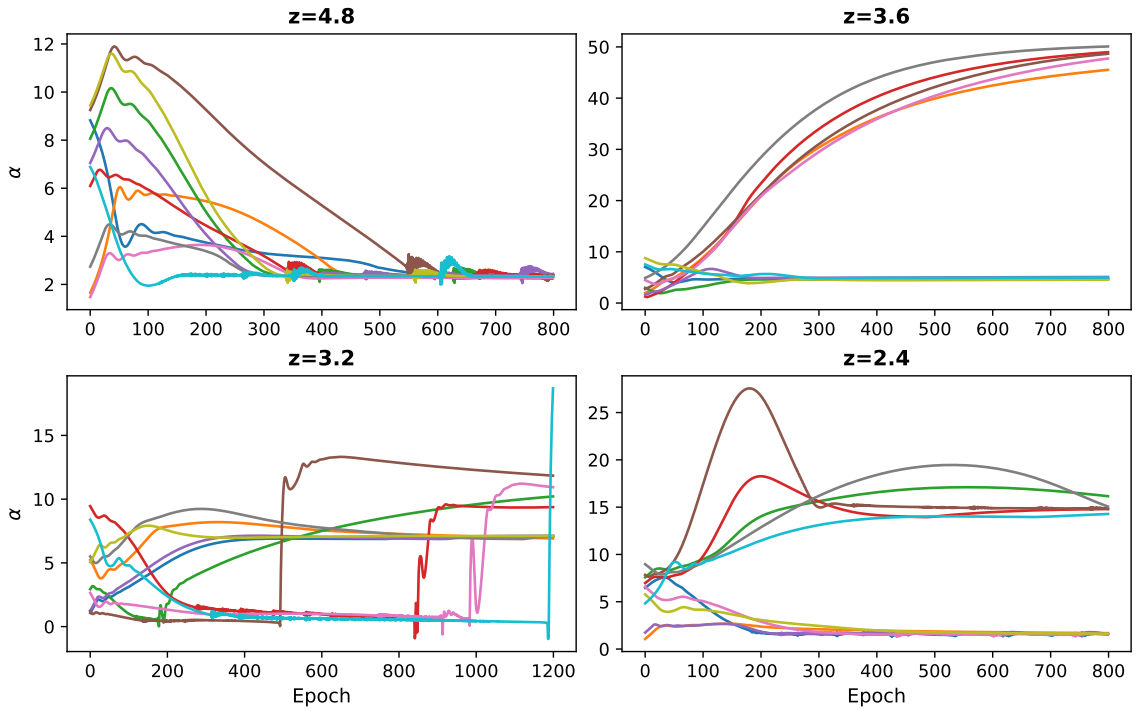


Figure 3.11: Parameter α through 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4 as a function of training epoch.

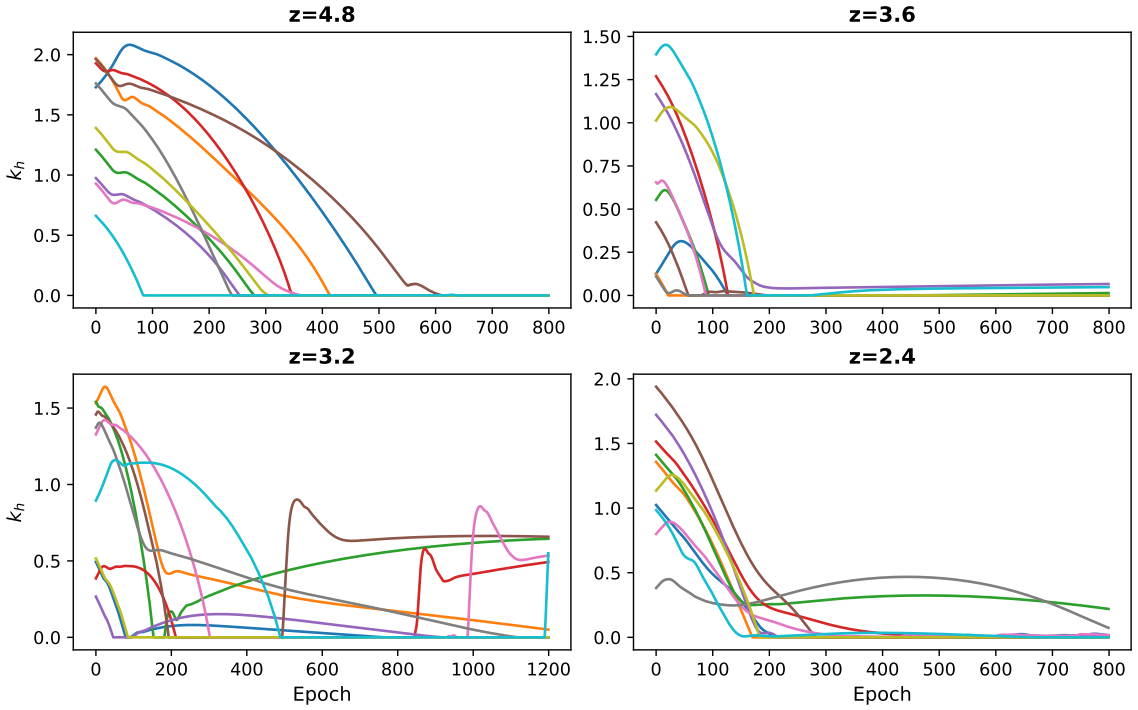


Figure 3.12: Parameter k_h through 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4 as a function of training epoch.

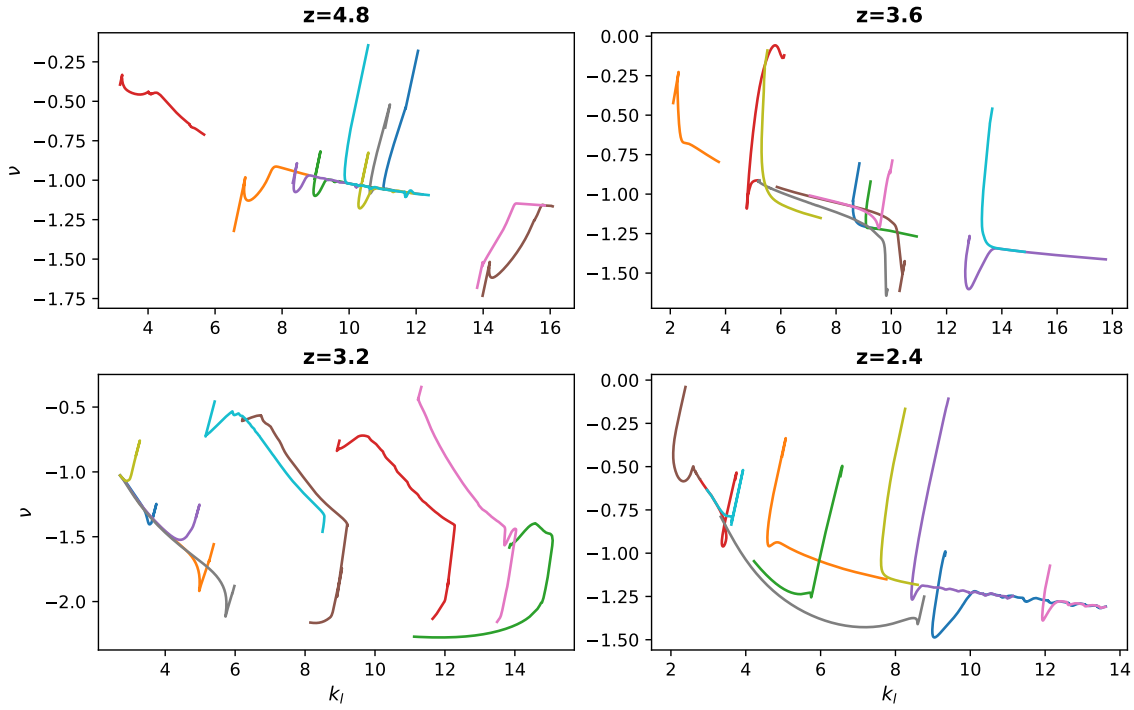


Figure 3.13: Parameter k_l (x axis) vs parameter ν (y axis) throughout 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4.

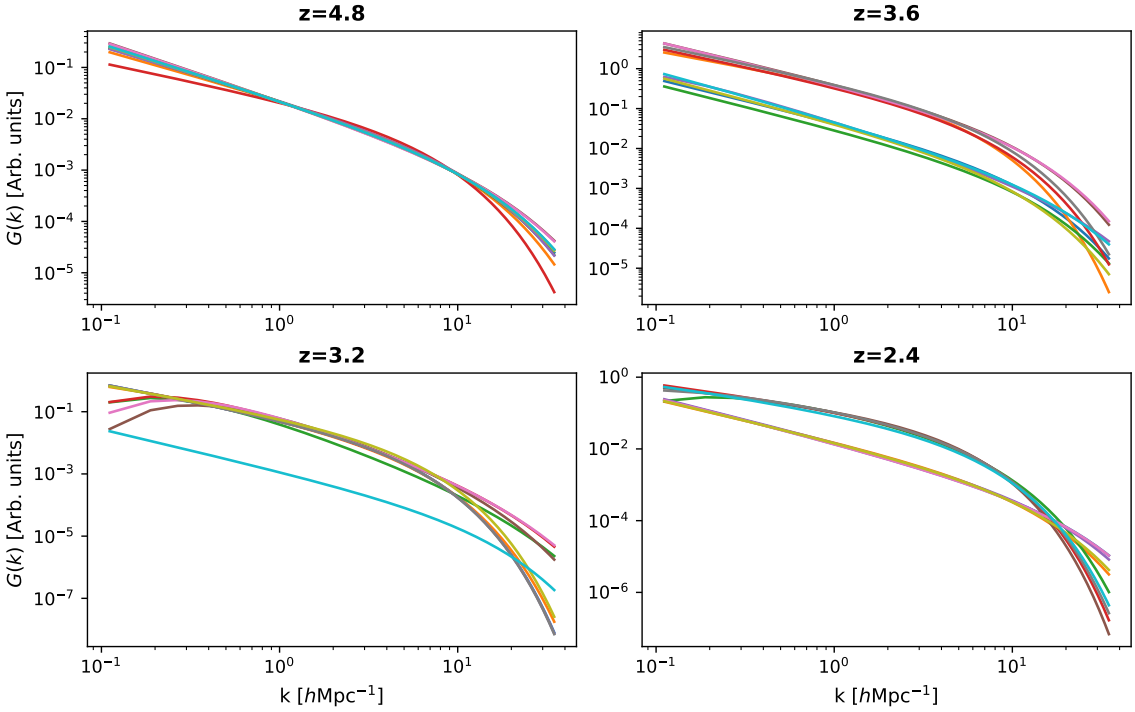


Figure 3.14: Green's functions at best validation loss for 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4.

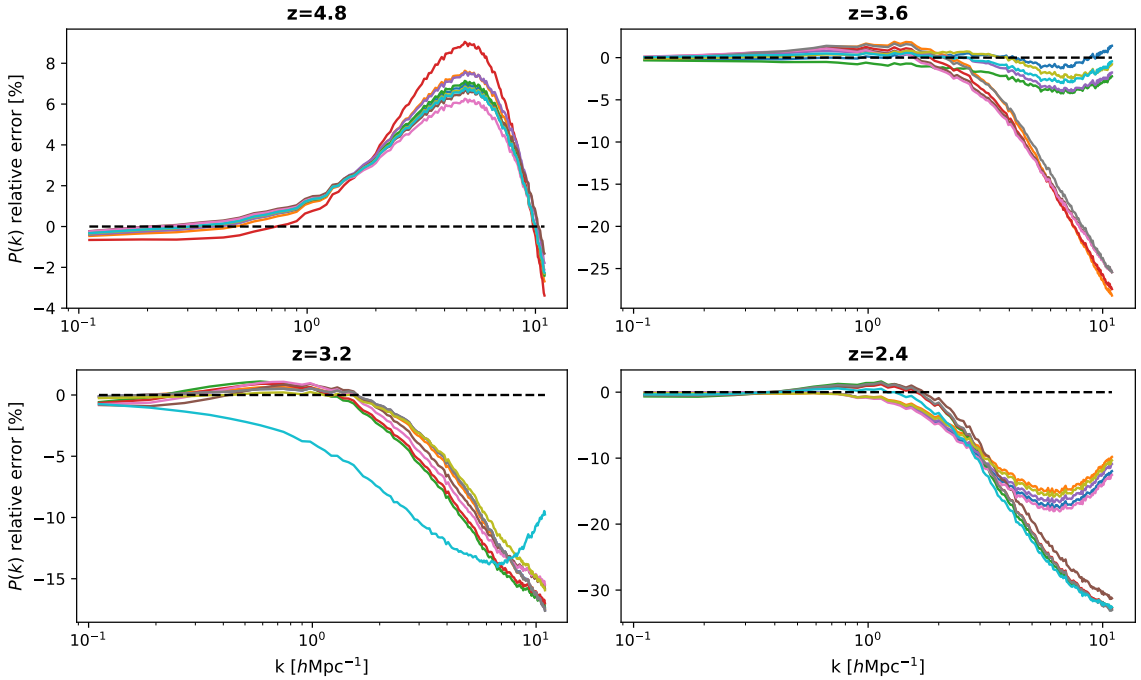


Figure 3.15: Power spectra relative differences with respect to the Sherwood reference for 10 optimizations per redshift for $z = 4.8$, 3.6, 3.2, and 2.4. The parameters used for the `JERALD` output maps are the ones associated to the best validation losses for every optimization.

3.3 Cosmological and astrophysical inputs

Even in the comparatively benign setting of the analysis conducted so far, where a single, continuous, one-dimensional quantity, z, was varied, the recovered parameters prove poorly behaved, revealing the existence of multiple minima whose relative depth changes with redshift, near-flat directions along which the loss barely varies, and an optimum that relocates discontinuously rather than tracking a smooth trajectory. The question this section turns to is whether these pathologies persist once the quantity being varied is the set of cosmological and astrophysical parameters $\boldsymbol{\phi}$ themselves, which is the case of actual interest for an emulator, and, if so, what form they take and whether anything can be done about them.

All the investigations that follow use the CAMELS simulation suite [182], with the IllustrisTNG [183, 184] physics prescription, as reference. Two of its sets are employed: the Latin hypercube (LH) set, comprising 1000 simulations sampling Ω_m and σ_8 together with four astrophysical parameters $\{A_{SN1}, A_{AGN1}, A_{SN2}, A_{AGN2}\}$ controlling the strength and efficiency of supernovae and AGN feedback, each realized with a different random seed; and the cosmic variance (CV) set, in which all physical parameters are held at their fiducial values ($\Omega_m = 0.3$, $\sigma_8 = 0.8$, $A_{SN1} = A_{AGN1} = A_{SN2} = A_{AGN2} = 1$) and

only the seed of the initial conditions varies. The two therefore isolate, respectively, the dependence on $\boldsymbol{\phi}$ and the dependence on the realization alone. Both feature a comoving box of $L = 25h^{-1}$Mpc with 256^3 dark matter particles and gas elements, considerably smaller than the $80h^{-1}$Mpc Sherwood-Relics volumes used in Chapter 2. Target dark matter density fields are extracted at $z = 0$ on a 128^3 mesh.

The input fields for the model are generated at $z = 0$ with FastPM following the same procedure as in Section 2.3.2: initial conditions are produced at $z = 9$ with 2LPT from a linear density map of 256^3 cells, i.e. twice the resolution of the simulation mesh, in order to improve accuracy at small scales, and 128^3 particles are then evolved on a 128^3 mesh in 10 FastPM steps with force resolution $B = 2$, the output being painted onto a 128^3 mesh to match the target. The random number generator used by CAMELS to draw the initial white noise field is the same one used for the Sherwood-Relics simulations (part of N-GenIC [160]), and is compatible with that of FastPM. Therefore, as in Chapter 2, feeding FastPM the seed of a given CAMELS simulation together with the corresponding linear power spectrum yields an approximate realization of the same structures as its full-hydrodynamical counterpart.

All the optimizations reported in the following sections were performed on the Leonardo supercomputer in CINECA, a single one requiring approximately 0.25 seconds per step and roughly 13 minutes for all 3000 steps at the resolution described above; the analyses of the following sections are based on ~ 5000 such runs in total.

3.3.1 Simplified model and training setup

Before turning to the results, it is important to mention that the analyses that follow differ from those of the preceding sections in two respects, both introduced here and retained throughout the remainder of the chapter: the removal of one model parameter, and the use of the full simulation volume for training.

Removal of the large-scale cutoff

The Green's function of Eq. (2.7) contains two exponential factors, one suppressing the displacement below k_h and one above k_l. As is already apparent from the redshift sweep of Section 3.2 (Figure 3.12), at least at higher redshift, as well as noticed in Chapter 2, in the optimizations discussed so far k_h was consistently recovered at or very near zero, i.e. the large-scale suppression was effectively switched off by the optimization itself. Irrespectively of the reason behind this mechanism, it appears that this parameter, in most cases, brings no meaningful contribution: it was therefore removed, so that the

Green's function used throughout the following sections is

$$\tilde{G}(k) = \alpha\, e^{-k/k_l}\, k^{\nu} \quad , \tag{3.2}$$

leaving four free parameters, γ, α, k_l and ν, for a single displacement layer. This choice has the benefit that the degeneracy structure of the model becomes considerably easier to characterize, and the analysis of Section 3.3.3 is correspondingly cleaner; second, four parameters is few enough that direct visualization of the loss landscape becomes somewhat feasible, an avenue not pursued here, but which the behavior documented below suggests would be informative. Section 3.4.2 also verifies that this simplification does not introduce spurious effects in the following analysis as when k_h is reinstated, none of the conclusions change, at least as far as the CV results are concerned.

Training on the full volume

The second change concerns the training region. Both in Chapter 2 and in the previous sections the model was trained on a subset of the box, with a held-out region used for validation and for selecting the parameters to retain. In what follows, the loss of Eq. (2.10) is instead evaluated over the entire simulation volume, and the parameters retained are those achieving the lowest training loss. Three considerations motivate this. First, the usual reason for holding out a validation region, i.e. guarding against overfitting, has little weight here: the model has four free parameters, against $\sim 2 \times 10^6$ voxels for a 128^3 mesh; the ratio of model capacity to data is such that memorizing the target field is not a possibility, in contrast to the neural emulators discussed in Section 1.4.2, for which validation is indispensable. Second, this expectation was verified directly. Optimizations were run on a number of simulations in the CAMELS CV dataset both ways, i.e. training on the full box and on a sub-box with a held-out validation region, and the resulting parameters were then applied to different realizations with the same $\boldsymbol{\phi}$. The two procedures generalize equally well; if anything, parameters obtained from the full box perform marginally better, as would be expected from the larger effective training set (as an example, see Figure 3.16). Third, training on the full periodic box is what makes the argument of Appendix E available: the derivation there applies Parseval's theorem to the loss summed over the entire volume, which is what allows to identify how γ and $J(k)$ enter in it. Restricting the sum to a sub-volume would amount to introducing a window function, whose transform would convolve the spectra appearing in the derivation and obscure that identification. The choice of training region is therefore not merely convenient, but is what makes the loss analytically tractable.

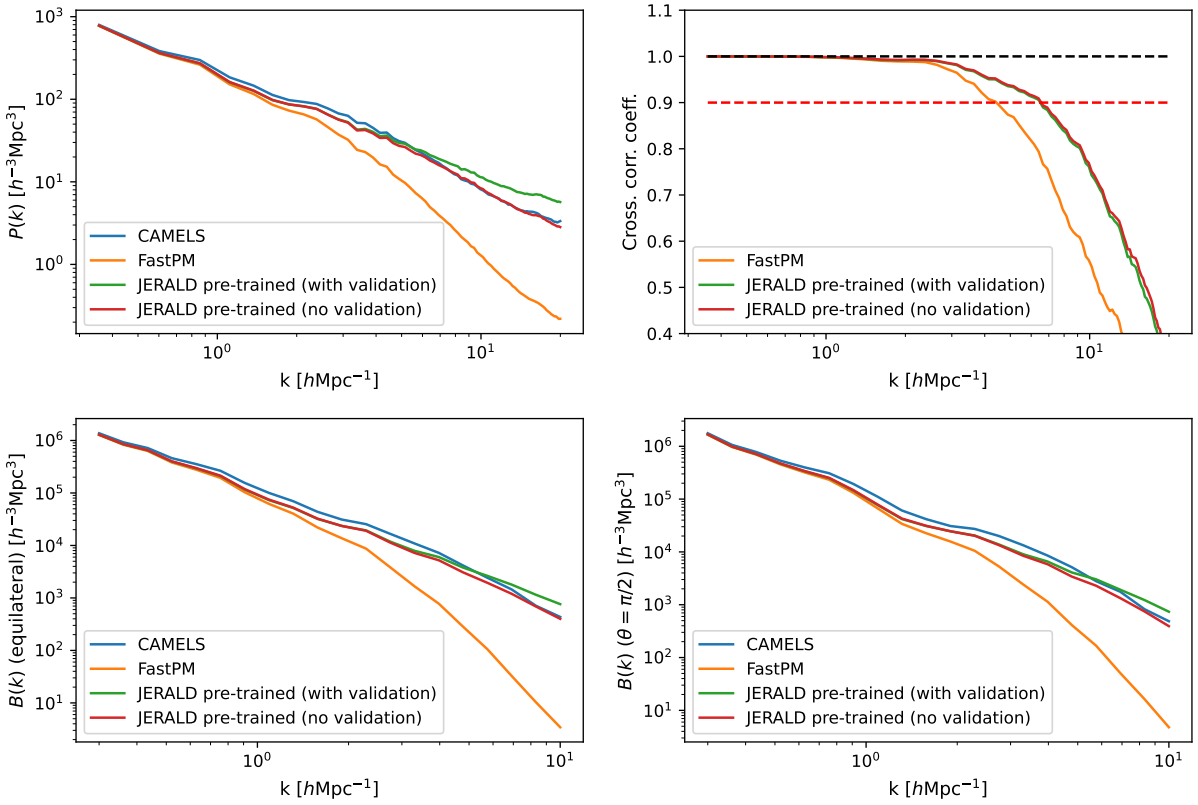


Figure 3.16: Performance of JERALD on the CAMELS CV14 simulation with parameters optimized on the CAMELS CV2 simulation with and without validation loss. *Top:* power spectra (left) and cross correlation coefficients (right), *bottom:* bispectra for varying arm lengths $k_1 = k_2$ for an equilateral triangle configuration (left) and for $\theta = \pi/2$ (right). In every case the generalization performance of JERALD optimized with validation loss (green lines) is almost the same as that without validation loss (red lines).

3.3.2 Direct regression using a neural network

Despite the results of Section 3.2, the most direct way to find a relation $\boldsymbol{\theta}(\boldsymbol{\phi})$ is the one that was attempted first. JERALD was optimized on 700 simulations from the LH dataset, with the intention of then fitting a neural network to predict the optimized parameters $\boldsymbol{\theta}$ from the corresponding $\boldsymbol{\phi}$ of each simulation[1]. Several optimization schemes were tested in order to make the mapping retrievable: multiple different random initial guesses per simulation, in order to allow optimizations to find the best optima; a shared initial guess reused for every LH simulation, in order to potentially converge to similar optima; and increasing the number of optimization steps, to allow a more thorough exploration of the loss landscape. In every configuration the outcome was the same. The optimized parameters showed no discernible structure as a function of $\boldsymbol{\phi}$, and the network was unable to predict them beyond reproducing their sample means; specifically, models flexible enough

[1] Note that this regression is different from the approach in Section 3.1.1: there, the NN is integrated directly into the model and its weights become the objects to optimize; here, on the contrary, the model still optimizes $\boldsymbol{\theta}$, and the NN is only used to find the mapping $\boldsymbol{\theta} \mapsto \boldsymbol{\phi}$.

to fit the training simulations failed to generalize to held-out ones, while more constrained models converged to a $\boldsymbol{\phi}$-independent prediction. One possible explanation is that the optimum reached by `JERALD` depends less on $\boldsymbol{\phi}$ than on the specific realization of the input FastPM field, which the network never sees; this is especially plausible given the small $25h^{-1}$Mpc CAMELS boxes, in which large-scale modes are sparsely sampled. Supplying the network with information about the input simulation directly, in the form of either its power spectrum or bispectrum alongside $\boldsymbol{\phi}$, produced only a marginal improvement, although still nothing substantially better than the mean.

This, however, highlights an additional difficulty that the redshift analysis above ignored: there, every snapshot came from the same simulation, so whatever cosmic variance was present was common to the entire sequence; across the LH suite this is no longer true. Each simulation differs from every other not only in its physical parameters but also, importantly, on its initial conditions, so the scatter in the recovered $\boldsymbol{\theta}$ mixes any dependence on $\boldsymbol{\phi}$ with a dependence on the particular realization, and no amount of inspection of the suite alone can separate the two, since the design provides no repeated draws at fixed $\boldsymbol{\phi}$. Fitting a regression to such a sample is therefore attempting to isolate a signal without any measurement of the noise it sits on. The situation is aggravated by resolution: the CAMELS boxes used here feature $L = 25h^{-1}$Mpc, against the $80h^{-1}$Mpc Sherwood volumes of the the redshift analysis, and small boxes sample large-scale modes poorly, inflating exactly the realization-to-realization fluctuations at issue, as will be discussed also in Chapter 4.

For this reason, in the following sections, the two effects are addressed separately. Specifically, Section 3.3.3 quantifies how much the optimized parameters vary at fixed $\boldsymbol{\phi}$, using the CAMELS CV dataset, establishing the level of scatter attributable to cosmic variance and to the optimization itself, independently of any parameter dependence, and attempting to identify the origin of the degeneracies highlighted in the previous sections and bypass them. From sections 3.4 onward, the focus returns mostly to the full suite, in which $\boldsymbol{\phi}$ varies, asking whether a dependence on $\boldsymbol{\phi}$ can be recovered on top of that baseline using the insights derived from the CV analysis.

3.3.3 Fixed cosmology: cosmic variance and parameter degeneracies

A series of `JERALD` optimizations was run on the CAMELS CV set, comprised of simulations sharing identical cosmological and astrophysical parameters but generated from different random seeds for the initial conditions. Eleven CV realizations were considered, with 35 independent optimizations of 3000 steps, each starting from a different randomly drawn set of initial guesses; a deeper run of 200 optimizations was performed on a single

realization (CV2) to characterize the parameter distribution more finely. The resulting parameters, selected by best training loss, are shown in Figure 3.17.

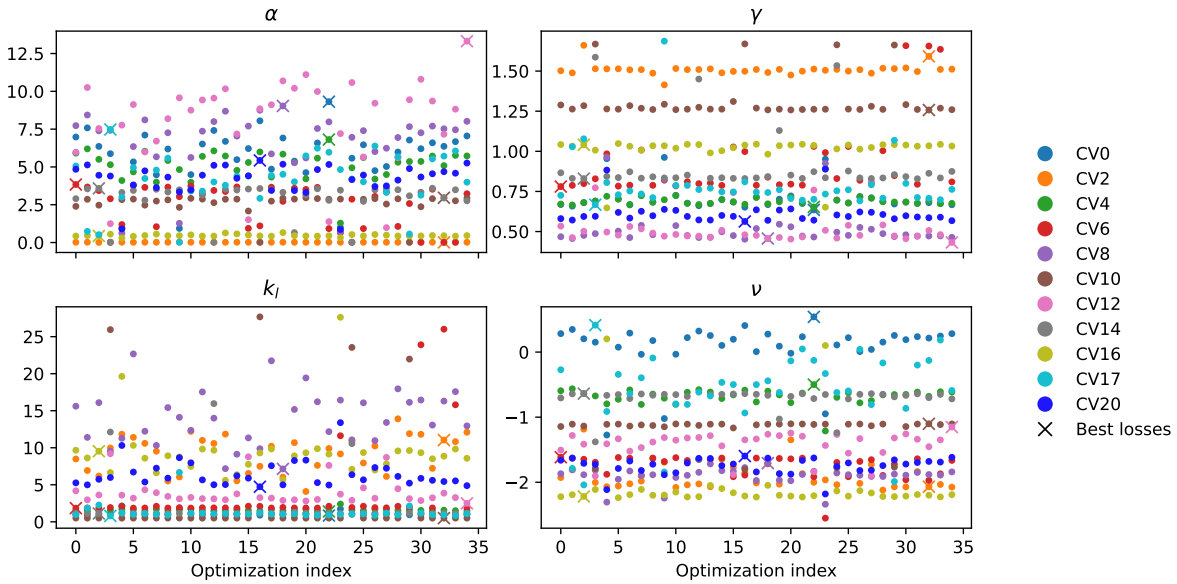


Figure 3.17: Optimized `JERALD` parameters for 11 different CAMELS CV simulations. The x axis simply represents an index of the optimization (35 optimizations with different initial guesses for $\boldsymbol{\theta}$ per CV index). The crosses highlight the parameters associated to the best losses across all 35 optimizations with fixed CV index.

Figure 3.17 exhibits the two distinct problems, which should be considered separately. First, the recovered parameters differ substantially between CV realizations, despite those realizations sharing identical $\boldsymbol{\phi}$: this variation, hypothesized above, cannot reflect any dependence on cosmology or astrophysics, and must instead be attributed to the realization of the initial conditions. Second, even within a single realization with fixed $\boldsymbol{\phi}$ and fixed initial conditions, several of the parameters remain widely scattered across the 35 optimizations. Either problem alone would be enough to defeat a regression of $\boldsymbol{\theta}$ on $\boldsymbol{\phi}$: the first because it introduces variance uncorrelated with the regressor, the second because it means $\boldsymbol{\theta}$ is not even a well-defined function of the simulation being fitted. The second issue was anticipated in Chapter 2 and Section 3.2: some parameters are degenerate by virtue of how they enter the model, so that different values describe the same physical correction. Figure 3.18 makes this explicit for k_l and ν in the CV2 case, where the 200 optimizations do not cluster around a point but rather they populate a curve of nearly constant loss. As observed in Section 3.2, its origin is apparent from the Green's function: with k_h removed, Eq. (2.7) reduces to

$$\tilde{G}(k) = \alpha\, e^{-k/k_l}\, k^{\nu} \quad ,$$

so that $\log \tilde{G}(k) = \log \alpha - k/k_l + \nu \log k$. The two parameters therefore contribute a

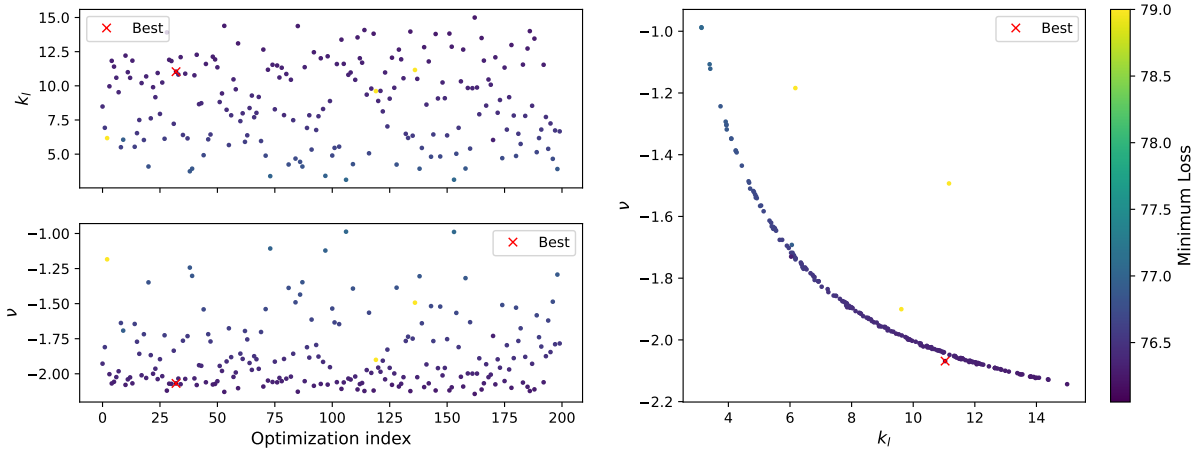


Figure 3.18: Optimized JERALD parameters k_l and ν for 200 optimizations on CV2 starting from different initial random guesses colored by the minimum loss achieved, showing their degeneracy.

term linear in k and a term linear in $\log k$ respectively, and over the limited range of wavenumbers resolved by the target simulation a change in one can be compensated by a change in the other. Figure 3.19 confirms this neatly: the widely dispersed parameter values of Figure 3.18 correspond to Green's functions that are nearly indistinguishable over the relevant scales. This suggests the appropriate object of study is not the parameter vector but $\tilde{G}(k)$ itself, which is invariant under the degeneracy.

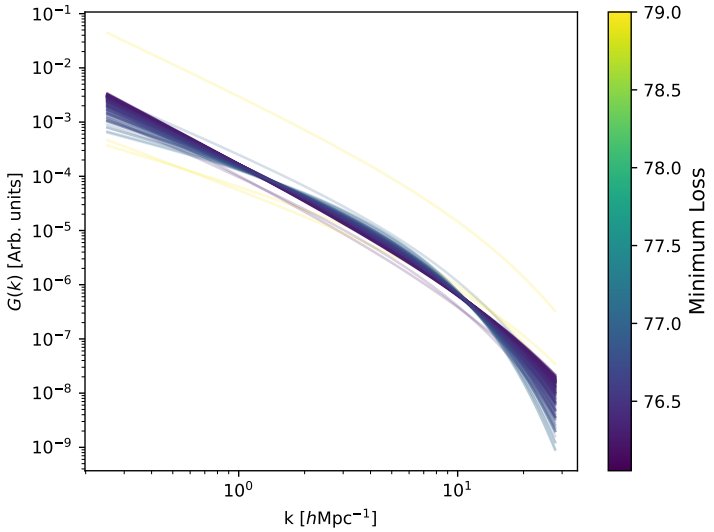


Figure 3.19: Green's functions for the optimized JERALD parameters in Figure 3.18.

In a first test to exploit this finding, a neural network was trained on the 700 LH simulations of Section 3.3.2 to predict $\tilde{G}(k)$ rather than the parameters directly, with γ handled separately as it does not enter the Green's function. The training objective was correspondingly split, combining a mean-squared error on $\tilde{G}(k)$ with a second term

on γ, weighted against each other by a free hyperparameter. This, however, was not investigated deeply for two reasons: other lines of investigation, described below, proved more promising in the meantime; and the split objective is unsatisfactory in itself, since the two terms carry different units and dynamic ranges, making the relative weight between them arbitrary rather than physically motivated.

Dependence on γ

Focusing on $\tilde{G}(k)$ rather than the parameters that define it resolves the within-realization scatter, but not the across-realization one. As the left panel of Figure 3.20 shows, the Green's functions recovered at fixed CV index cluster tightly, yet the clusters themselves sit at visibly different positions for different realizations. The same is true for γ, which is not absorbed by the Green's function and continues to vary appreciably between CV indices, although clustering around fixed values for fixed CV (Figure 3.17)[2].

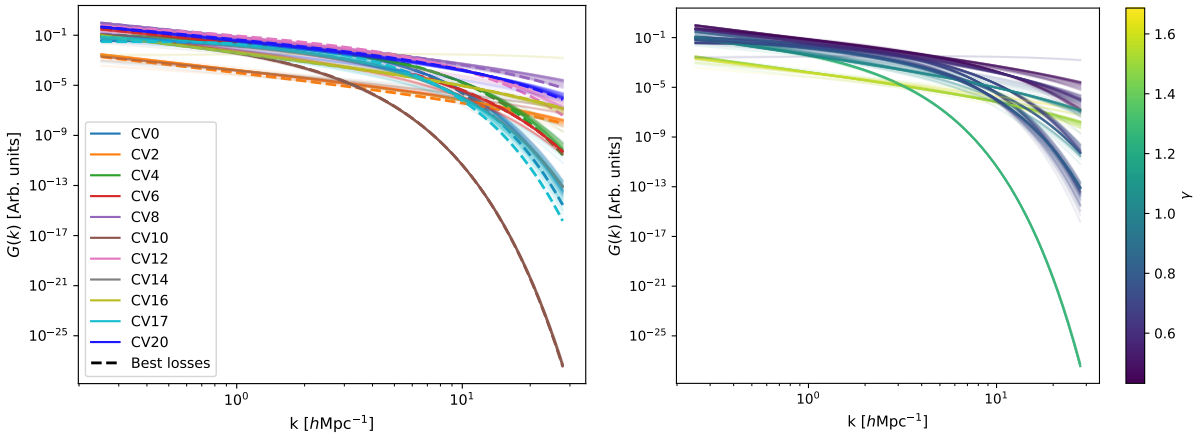


Figure 3.20: Optimized Green's functions for all CV optimizations considered. *Left:* colored by CV index, showing the best loss for fixed CV index as a dashed line, *right:* colored by associated γ.

However, the right panel of Figure 3.20, in which the same curves are colored by their associated γ, shows that the two residual variations are not independent: the vertical offset of $\tilde{G}(k)$ tracks γ systematically. Given the definition of the model, this is expected: the displacement of Eq. (2.2) is generated by the product $\tilde{G}(k)\,\tilde{f}(\boldsymbol{k})$, with $f = \rho^\gamma$, where ρ is the input (normalized) DM density, so neither factor is separately constrained by the loss, only their combination is. Since the typical amplitude of the source term at a given scale is measured by $\sqrt{P_{\rho^\gamma}(k)}$, $P_{\rho^\gamma}(k)$ being the power spectrum of the source $f = \rho^\gamma$, the natural scale-dependent measure of the correction the model actually applies is obtained by multiplying the optimized Green's function by that quantity, and a compensating shift

[2] The apparent tendency of γ to cluster near values of the form $i/4$, with i integer, is set aside here; it is briefly returned to in Section 3.5.1.

in γ and in the normalization of $\tilde{G}(k)$ should then leave the product unchanged. Figures 3.21 and 3.22 confirm this: the spread of the resulting curves across CV indices is reduced considerably with respect to that of $\tilde{G}(k)$ alone, at large and intermediate scales.

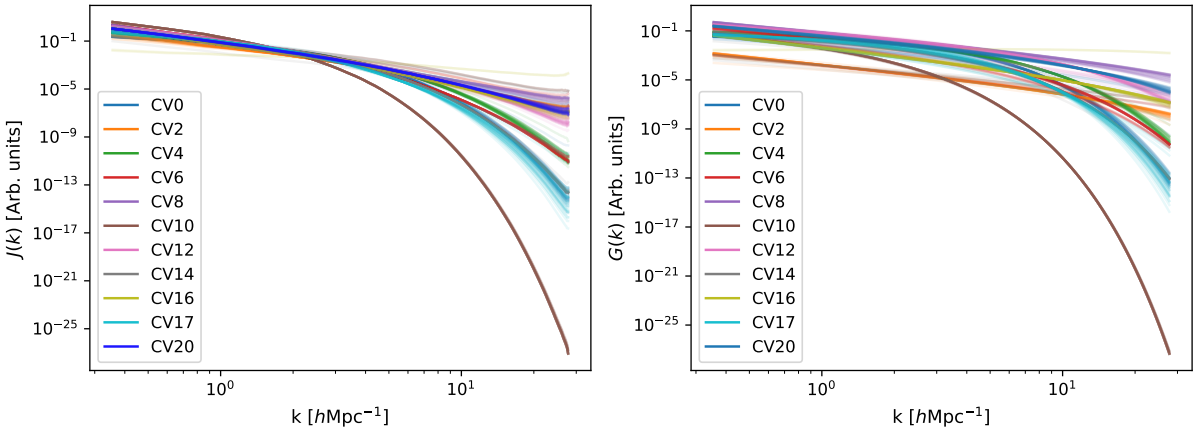


Figure 3.21: Optimized Green's functions multiplied by $\sqrt{P_{\rho^\gamma}(k)}$, the square root of the power spectrum of the source field ρ^γ (*left*), compared with the Green's functions alone (*right*), for all CV optimizations.

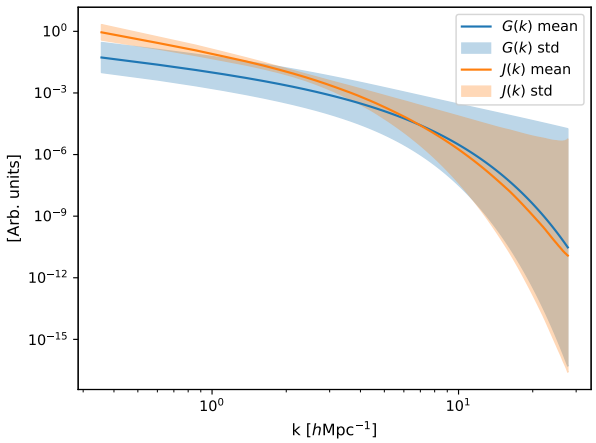


Figure 3.22: Same as figure 3.21, showing the average curves with the associated spread (1σ, in log space).

This indicates that the quantity

$$J(k) = \tilde{G}(k)\sqrt{P_{\rho^\gamma}(k)} \tag{3.3}$$

is well constrained at fixed cosmological and astrophysical parameters, and therefore largely insensitive to the realization of the initial conditions, at least at large and intermediate scales, even though the individual parameters entering it are not. A network trained with a loss on $J(k)$ would then predict a single object rather than a set of separately weighted ones, removing the arbitrary hyperparameter of the split objective

discussed above. Appendix E shows formally how $J(k)$ emerges as the scale-dependent object that `JERALD` optimizes, while any phase dependence is delegated to γ, allowing it to be determined consistently across optimizations at fixed realization (Figure 3.17).

A cheap approximation for the source power spectrum

Training a network to predict $\boldsymbol{\theta}$ with a loss on $J(k)$ is computationally demanding, since every example in the training set requires the power spectrum of the corresponding source field ρ^γ, and γ changes as the network is updated, so the quantity cannot simply be precomputed once and reused across epochs[3]. This cost would be avoided if $P_{\rho^\gamma}(k)$ could be approximated by an expression involving only the power spectrum of the input field $P_\rho(k)$, which is independent of γ and can therefore be evaluated once per simulation and stored. The natural candidate is a relation between $P_{\rho^\gamma}(k)$ and $P_\rho(k)^\gamma$.

The ratio of $\sqrt{P_\rho(k)^\gamma}$ to $\sqrt{P_{\rho^\gamma}(k)}$, shown in the left panel of Figure 3.23, is not constant but retains a residual dependence on k. This dependence is well absorbed by a single power of the wavenumber, leading to the approximation

$$J'(k) = A_\gamma \, \tilde{G}(k) \sqrt{P_\rho(k)^\gamma \, k^{3(\gamma-1)}} \quad , \tag{3.4}$$

where the exponent $3(\gamma-1)$ is not fitted but is the scaling predicted by the argument discussed in the next section, and the normalization A_γ was fitted numerically as

$$A_\gamma \simeq \exp\!\big[-0.95 + \beta \, \log\!\big(1 + e^{\,\kappa/\beta\,(\gamma-\gamma_0)}\big)\big] \ , \quad \beta \simeq 0.742 \ , \quad \kappa \simeq 3.22 \ , \quad \gamma_0 \simeq 0.78 \ . \tag{3.5}$$

This fit was performed on the log difference between $\sqrt{P_\rho(k)^\gamma \, k^{3(\gamma-1)}}$ and $\sqrt{P_{\rho^\gamma}(k)}$ from the 500 optimizations on the LH dataset described in Section 3.3.4. Theoretically, the derivation for the exponent $3(\gamma-1)$ is valid mostly at large k; nonetheless, the argument is heuristic, thus the log difference was computed at $k \simeq 2h\mathrm{Mpc}^{-1}$, in order to avoid noise from poorly-sampled modes at small k and small-scale effects at large k. The bracketed term in Eq. (3.5) is a softplus in γ, so that A_γ saturates at $\sim e^{-0.95}$ for $\gamma \ll \gamma_0$ and grows as $\exp[\kappa(\gamma-\gamma_0) - 0.95]$ for $\gamma \gg \gamma_0$. With this choice the ratio becomes close to scale-independent over the range of wavenumbers of interest, as shown in the right panel of Figure 3.23.

Finally, preliminary results indicate that much of the γ-dependence absorbed by $J(k)$ is carried by its normalization rather than its scale dependence, which, if confirmed, would permit a considerably cheaper approximation. This observation was not investigated in detail and would need to be established before being relied upon.

[3] A parallelized and differentiable implementation of the power spectrum, suitable for use directly inside a loss, is available in the `JERALD` code and is described in Appendix C; it has not yet been employed for this purpose.

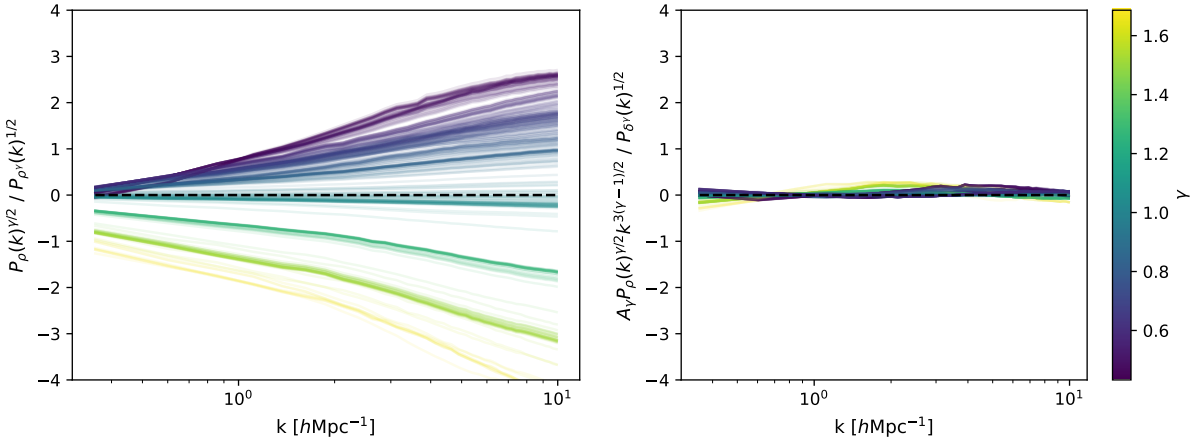


Figure 3.23: Ratio between $\sqrt{P_\rho(k)^\gamma}$ and $\sqrt{P_{\rho^\gamma}(k)}$ across all CV optimizations, colored by γ, before (*left*) and after (*right*) including the correction factor $A_\gamma\, k^{3(\gamma-1)/2}$ of Eq. (3.4).

Scaling of the source power spectrum

The exponent $3(\gamma-1)$ identified in the previous section is the result of an approximate scaling argument. Let δ be a Gaussian random field with power spectrum $P(k)$ and consider δ^n for integer $n \geq 1$. Since a product in configuration space is a convolution in Fourier space,

$$\widetilde{\delta^n}(\boldsymbol{k}) = \int \prod_{i=1}^{n-1} \frac{d^3 q_i}{(2\pi)^3}\, \tilde{\delta}(\boldsymbol{q}_1)\cdots\tilde{\delta}(\boldsymbol{q}_{n-1})\, \tilde{\delta}\Big(\boldsymbol{k} - \textstyle\sum_i \boldsymbol{q}_i\Big) \quad , \tag{3.6}$$

so that P_{δ^n}, the correlation between $\widetilde{\delta^n}(\boldsymbol{k})$ and $(\widetilde{\delta^n})^*(\boldsymbol{k}')$, involves the $2n$-point function of a Gaussian field. Wick's theorem decomposes this expression into $(2n-1)!!$ products of 2-point correlations; considering only the maximally coupled terms, in which each contraction joins a leg of $\widetilde{\delta^n}(\boldsymbol{k})$ to one of $\widetilde{\delta^n}(\boldsymbol{k}')$, the contribution to P_{δ^n} is easily shown to be (using Eq. (1.27)):

$$P^c_{\delta^n}(k) \propto \int \prod_{i=1}^{n-1} \frac{d^3 q_i}{(2\pi)^3}\, P(q_1)\cdots P(q_{n-1})\, P\Big(\Big|\boldsymbol{k} - \textstyle\sum_i \boldsymbol{q}_i\Big|\Big) \quad . \tag{3.7}$$

Each self contraction in the remaining terms of the Wick expansion trades a k-dependent propagator for a constant $\langle\delta^2\rangle$. The scaling of each class then follows by dimensional analysis. Taking $P(k) \propto k^\eta$ over the range of wavenumbers of interest and assuming the convolution integrals to be dominated by $q \sim k$, a term with m propagators $C_m(k)$ carries $(m-1)$ integration measures and m factors of the power spectrum, so that

$$C_m(k) \;\propto\; k^{m\eta+3(m-1)} \;\propto\; P(k)^m\, k^{3(m-1)} \quad . \tag{3.8}$$

Since two such terms differ in their exponent by $(n-m)(\eta+3)$, and the effective slope of the FastPM spectrum is $\eta \simeq -2$ at the scales of interest, the maximally coupled term $m = n$ dominates at large k, whereas the $m = 1$ term, simply proportional to $P(k)$, dominates at small k. Retaining the former,

$$P_{\delta^n}(k) \;\propto\; P(k)^n \, k^{3(n-1)} \quad . \tag{3.9}$$

The assumption that the integrals are dominated by $q \sim k$ involve requiring that, after rescaling $\boldsymbol{q}_i = k\boldsymbol{x}_i$ in Eq. 3.7, the remaining dimensionless integral converge, which is verified when $-3 < \eta < -3/2$, a window comfortably containing the slope considered. The argument remains heuristic in the present application, in that the FastPM field at low redshift is not Gaussian, the source term is $\rho^\gamma = (1+\delta)^\gamma$ rather than δ^γ, and γ is not an integer. It is used here only to fix the exponent $3(\gamma-1)$, with all remaining γ-dependent normalization are absorbed into the prefactor A_γ, and its adequacy is judged a posteriori from the right panel of Figure 3.23.

3.3.4 $J(k)$ as training objective

On top of highlighting the problems related to parameter degeneracies and identifying $J(k)$ as the combination the loss actually constrains, the analysis of the CV set also provides what Section 3.3.2 lacked, i.e. well-defined quantities to predict. It is therefore worth returning to the LH set and to the question left open there, but with $J(k)$ and γ in place of $\boldsymbol{\theta}$ as the targets of a regression. The other components of $\boldsymbol{\theta}$ are determined only up to transformations that leave the model's predictions unchanged, therefore a regressor would be attempting to reproduce scatter carrying no information, which is a plausible reading of why the naive attempt of Section 3.3.2 returned nothing better than the sample mean; on the other hand, $J(k)$ is seemingly reproducible across independent optimizations of the same simulation and largely invariant under those exact transformations, while γ is reproducible across different optimizations for a fixed realization.

Additionally, since $J(k) = \tilde{G}(k)\sqrt{P_{\rho^\gamma}(k)}$, the knowledge of $J(k)$ together with γ determines $\tilde{G}(k)$ directly. Predictions of $J(k)$ and of γ would therefore specify the model completely, bypassing α, k_l and ν altogether, and with them the degeneracies among the three. Accordingly, the purpose of the following sections is to first establish which properties of the input field γ responds to, and then regress onto $J(k)$ itself. To this end, `JERALD` was optimized again on 500 simulations of the LH set, with three independent optimizations per simulation, and the lowest-loss parameters retained in each case; the next sections will all use this set of optimizations. Note that this set of optimizations is different with respect to the one in Section 3.3.2.

3.4 Predicting the model features

The scheme identified at the end of the previous section, i.e. predicting γ and $J(k)$, recovering $\tilde{G}(k)$ from the two, and evaluating the model without ever determining α, k_l and ν, reduces the emulation problem to two regressions. This and the next sections ask whether either can be carried out. Even if there is a practical reason for this analysis, i.e. a `JERALD` that can be applied to a new simulation without a new optimization, as mentioned at the beginning of this chapter, there is also an interpretive one: the model is physically motivated and its parameters were introduced with meanings attached; therefore, the fitted values should correlate with the properties of the field being corrected in ways that can be identified, and establishing what they do and do not respond to is informative regardless of whether a working emulator results.

It should be noted that the investigation reported here and in Section 3.6 were carried out concurrently rather than in sequence. They are presented in this order because the first provides context to motivate the second.

3.4.1 A physical hypothesis for γ

As noted in Section 3.3.3, γ is recovered consistently, up to numerical scatter, across independent optimizations of the same simulation, yet varies appreciably between realizations sharing identical cosmological and astrophysical parameters: this indicates that some properties of the input field itself, rather than in $\boldsymbol{\phi}$, appear to determine it. There is a physical argument for expecting this: the two ingredients of a displacement layer (Eqs. (2.2) and (2.7)) split `JERALD`'s job in a natural way: the Green's function $\tilde{G}(k)$ acts as a filter in scale, selecting the range of wavenumbers over which the correction operates, while the source term $f = \rho^\gamma$ reweights the relative contribution of overdense and underdense regions according to the value of the field rather than its scale. Between them, the two have considerable flexibility to absorb the physics missing from the approximate simulation, one acting on where the correction is applied in Fourier space, the other on which structures drive it. For this reason, γ should be sensitive to how much non-linear structure the input realization actually contains: this is the reason the tests below compare γ against clustering statistics measured directly on the FastPM input, rather than against the nominal σ_8 of the simulation. Additionally, the power of a field couples its Fourier modes: as the scaling argument at the end of Section 3.3.3 makes explicit, the power spectrum of ρ^γ contains convolutions of $P_\rho(k)$ at all orders up to γ, which is precisely the structure generated by non-linear gravitational evolution itself. The source term is thus capable of mimicking mode coupling, with γ controlling how strongly so, making it the parameter most likely to carry information about the non-linear state of the input field.

3.4.2 Correlations with the input field

To investigate the hypothesis in the previous section, γ was compared against three statistics measured on the input FastPM field of each simulation: the variance of the field smoothed with a top-hat of radius R, $\hat{\sigma}_R$, the cosmological skewness S_3 and kurtosis S_4 of the smoothed field δ_R, defined as

$$S_3 = \left\langle \delta_R^3 \right\rangle / \left\langle \delta_R^2 \right\rangle^2 \quad \text{and} \quad S_4 = \left(\left\langle \delta_R^4 \right\rangle - 3 \left\langle \delta_R^2 \right\rangle^2 \right) / \left\langle \delta_R^2 \right\rangle^3 \quad . \tag{3.10}$$

Note that the hat is used to indicate that the amplitude $\hat{\sigma}_R$ is estimated, specifically by integrating the measured power spectrum of the field against the top-hat window (see Eq. 1.29), over the range of wavenumbers accessible to the simulation. This range is bounded below by the first shell-averaged bin of the estimator, i.e. $\sim 0.35\,h\mathrm{Mpc}^{-1}$, and above by the Nyquist frequency of the mesh. Since the window is peaked at $k \simeq 2/R$, the choice of R is therefore bracketed on both sides: for $R \gtrsim 6h^{-1}\mathrm{Mpc}$, the peak falls below the accessible range, so the estimate is dominated by $P(k)$ bins at lowest wavenumber, which average over only a handful of modes and are correspondingly noisy (and additionally acquire spurious structure as the zeros of the window pass through them), for $R \lesssim 1h^{-1}\mathrm{Mpc}$, on the other hand, the radius approaches a few mesh cells ($\Delta = L/N \simeq 0.2h^{-1}\mathrm{Mpc}$) and the estimate is affected by MAS smoothing rather than by the field. A value $R = 2h^{-1}\mathrm{Mpc}$ is adopted throughout, comfortably inside both limits. Figure 3.24 plots the recovered γ for the 500 simulations of the LH set against $\hat{\sigma}_2$ computed on the associated FastPM inputs.

No trend is obvious by eye in the scatter, but it appears switching to binned means, also shown in the figure. The weak positive relation is confirmed by a Pearson correlation of $\rho = 0.34$, with a 95% confidence interval of [0.26, 0.43] obtained by bootstrap over the simulations, in agreement with the analytic Fisher interval. Crucially, the correlation survives conditioning on the parameters of the simulation, with a partial correlation of 0.32 when controlling for all six components of $\boldsymbol{\phi}$, indicating that the dependence between γ and $\hat{\sigma}_2$ is not the result of the cosmology acting through other channel. The converse test is also informative: the correlation between γ and the input σ_8 of each simulation is weaker ($\rho = 0.18$) and it vanishes entirely once the realized amplitude $\hat{\sigma}_2$ is controlled for, indicating that σ_8 influences γ predominantly through the actual amplitude measured on the non-linear box. The same is true for the remaining parameters in $\boldsymbol{\phi}$, i.e. any correlation, if present, vanishes when controlling for $\hat{\sigma}_2$. The one parameter for which this is not the case is $\mathrm{A_{SN1}}$, the galactic wind energy, whose correlation with γ, $\rho \sim 0.13$, is unaffected by such conditioning. This is unsurprising, since $\hat{\sigma}_2$ is measured on the FastPM field, which contains no baryons and therefore cannot encode it, and it is consistent with the argument of Appendix E, in which γ depends on the alignment between the source

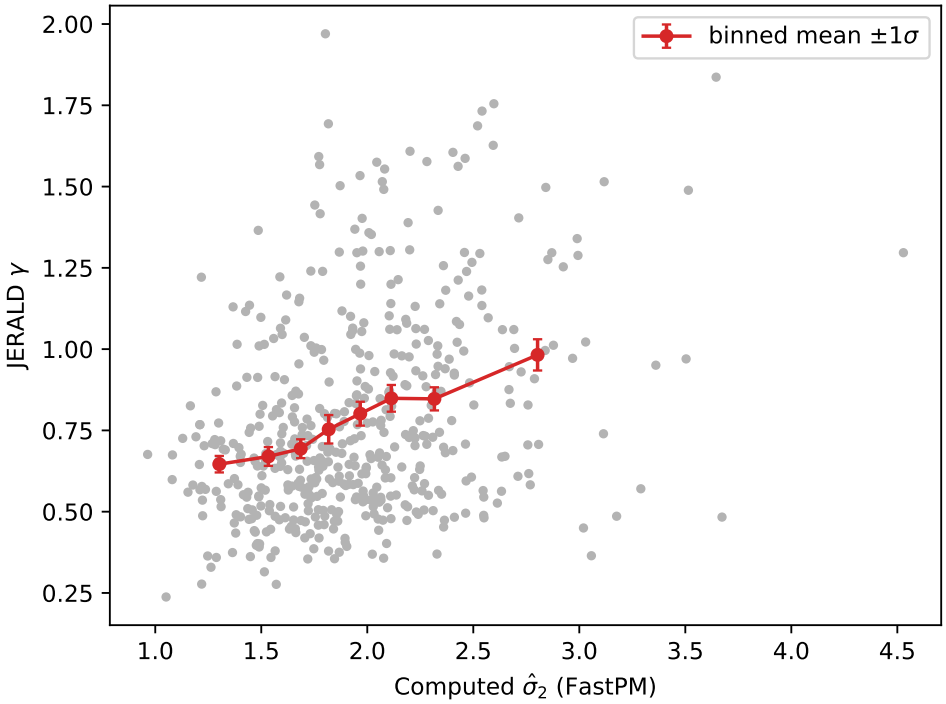


Figure 3.24: Scatter plot of γ optimized for 500 LH simulations against $\hat{\sigma}_2$ computed on the FastPM input. The red line was computed by binning both x and y axes into bins of (approximate) equal numbers of points.

term and the residual between reference and input, and as such it is not only influenced by the input field but also by the target. Nonetheless, the magnitude is under two per cent of the variance.

The dependence is, however, overwhelmingly on the realized amplitude $\hat{\sigma}_2$. Repeating the test with the reduced cumulants of the field S_3 and S_4 smoothed on the same scale gives correlations with γ of 0.15 and 0.11 with 95% confidence intervals [0.06, 0.24] and [0.02, 0.20] respectively, which fall to 0.10 and 0.09 once $\hat{\sigma}_2$ is controlled for, i.e. around one per cent of the variance in each case, against the eleven per cent accounted for by the amplitude alone. Since these quantities are constructed to be independent of the amplitude of the fluctuations, the conclusion is that γ responds predominantly to the clustering amplitude of the input realization.

The hypothesis of Section 3.4.1 is thus confirmed in its essentials, although in a measure that make it hardly useful for a predictor: $\rho^2 \simeq 0.11$ indicates that knowledge of $\hat{\sigma}_2$ accounts for around eleven per cent of the variance in γ. The dependence thus exists, and its origin is physically interpretable; however, it is not strong enough for γ to be predicted for a new, unseen simulation, as Section 3.5 will show.

Alternative parametrizations

The weakness of the dependence in the previous section and the residual scatter in γ may originate in the model rather than in the fields: $\boldsymbol{\theta}$ are recovered together only in combination, with multiple degenerate directions that contribute scatter to their optimized values, and large- and/or small-scale noise may leak into γ itself: for this reason, alternative parametrizations were briefly investigated, as one with fewer degenerate directions might yield a less noisy γ and a correspondingly sharper correlation. Four variants were optimized on the CV set with this in mind: in the first, k_l was removed, eliminating one of the near-flat directions; in the second, parameter k_h was reinstated, in order to check whether γ absorbs large-scale noise; in the third, the exponential cutoffs of Eq. (3.2) were replaced by Gaussian ones, $e^{-k^2/k_l^2} e^{-k_h^2/k^2}$, as in the original formulation of [149] (see Eq. (2.6)), in order to check whether the behavior of γ is tied to the assumed shape of the cutoff; in the fourth, the loss was restricted to the coarsest sub-bands of a discrete wavelet decomposition of both fields [185] (either the LLL band alone, or LLL together with the three bands containing a single high-pass direction), reducing the weight carried by the smallest scales, where both the approximate simulation and the argument of Appendix E are least reliable. Measuring the correlation itself requires the LH sample, and re-optimizing it under each variant is expensive; the only quantity accessible on the CV set is instead the spread of the recovered γ across realizations at fixed $\boldsymbol{\phi}$, which corresponds to the noise from which the correlation must be extracted. In none of the four variants was this spread appreciably reduced with respect to the baseline, nor was the clustering of $J(k)$ tighter or the performance of the optimized model materially affected. On that basis they were not carried through to the full LH set, and are reported here as avenues explored rather than as conclusive tests.

3.5 Direct regression of the model features

The tests of the previous section examined one summary statistic at a time, finding that γ is determined by the input field only weakly. However, that procedure is restrictive, since the input simulation contains much more information than any single number extracted from it; a natural next step is therefore to supply a regressor with the full power spectrum of the input field, together with the parameters of the simulation, and let it identify whatever combination is predictive. Additionally, the correlation analysis in the previous section addresses only the first of the two steps in the emulation scheme set out at the end of Section 3.3.4; the regressions in this section, on the other hand, address both.

Multiple input-output pairs were used: as input, either the power spectrum of the FastPM field truncated at a small-scale cutoff $k_{\max}$ (tested in the range $4 - 10\,h\mathrm{Mpc}^{-1}$), the cosmological and astrophysical parameters $\boldsymbol{\phi}$, or a concatenation of the two; as target,

either γ alone, or the full $\boldsymbol{\theta}$ vector with a loss defined on $J'(k < k_{\max})$ (see Eq. (3.4)), or directly $J(k < k_{\max})$ (not its approximation $J'(k < k_{\max})$, as in this case it does not need to be evaluated on the fly), bypassing the model parameters completely. Feed-forward neural networks, Gaussian process regression [186], gradient-boosted trees [187] and ridge regression [188] were used, with training and validation splits varying roughly around 400 of the 500 optimized LH simulations for training and 100 for validation. All inputs were standardized before training, using statistics computed on the training set alone. The components of $\boldsymbol{\phi}$ were standardized individually to zero mean and unit variance, after taking logarithms of the four feedback parameters, which are sampled logarithmically in the CAMELS LH dataset (see this link). Power spectra were used in logarithmic form, with the mean over the training set subtracted by k-bin, and rescaled to unit overall variance. Targets were standardized in the same way and with the same convention, using training-set statistics. It is important to highlight the fact that a validation set of ~ 100 simulations and effects to search for of the order of a few per cent of the variance (see Section 3.4.2), the outcome of any single fit is dominated by the particular train/validation split. Individual figures are therefore not meaningful, and everything quoted below is either a median over thirty independent splits, or a comparison between two configurations evaluated on the same split, the latter being more informative as the split-to-split variation cancels when evaluating differences on the same split. The results reported below refer to the following configurations of the four model classes. The feed-forward neural networks feature a single hidden layer of eight units, implemented in `JAX` [155] and trained with AdamW [175] as provided by `Optax` [189], with L_2 regularization of weight decay $\lambda = 10^{-2}$; other depths and widths and regularization strengths were tested, producing comparable results. Ridge and Gaussian process regression were used as implemented in `scikit-learn` [190]: for the former the regularization strength was selected by leave-one-out cross-validation over 25 logarithmically spaced values in $[10^{-3}, 10^{3}]$, performed within the training set alone so that the held-out simulations play no part in the choice; for the latter the kernel was taken as a constant-scaled squared-exponential plus a white-noise term, with hyperparameters obtained by maximizing the marginal likelihood and targets standardized internally. Gradient-boosted trees were implemented via `XGBoost` [191], with 300 estimators of maximum depth 3, learning rate 0.05 and L_2 leaf regularization $\lambda = 1$.

3.5.1 Predicting γ

Ridge regression on the single quantity $\hat{\sigma}_2$ accounts for a median $R^2 \simeq 0.10$ of the variance in γ on held-out simulations: this reproduces the value obtained in Section 3.4.2 from the Pearson correlation between the two quantities, which amounted to $\rho^2 \simeq 0.11$; since the two tests are entirely different, their agreement confirms that this dependence is both real and correctly quantified. Supplying the full power spectrum rather than a single

amplitude raises the median of the ridge regression to $R^2 \simeq 0.14$, the paired difference between the full power spectrum and $\hat{\sigma}_2$ alone being positive on the large majority of splits with a median gain of $\sim 4\%$; this seems to indicate that there is additional information in the scale dependence of the input spectrum beyond what any single smoothing radius captures, even though the increment is modest. Increased flexibility with respect to a linear model does not add extra information: Gaussian process regression on the full power spectrum input performs indistinguishably from ridge regression, and the same is true for the neural networks tested. Gradient-boosted trees perform worse, which is expected of an ensemble of piecewise-constant learners applied to a smooth and nearly linear function of many strongly correlated inputs at this sample size. These results are also insensitive to the small-scale cutoff imposed on the input: raising $k_{\max}$ from $4\,h\mathrm{Mpc}^{-1}$ to $10\,h\mathrm{Mpc}^{-1}$ leaves the predictability of γ unchanged within the split-to-split scatter. This indicates that γ mostly responds to features at large and intermediate scales, consistent with $\hat{\sigma}_2$ alone already accounting for most of what any input configuration achieves.

Concatenating $\boldsymbol{\phi}$ to the input yields only a small further improvement: the paired difference is positive on most of the splits, with a median gain of under 3% of the variance. This is consistent with the findings of Section 3.4.2: there, the partial correlations established that Ω_m and σ_8 carry practically nothing about γ once the realized clustering amplitude is known, while the galactic wind energy $\mathrm{A_{SN1}}$ retains a dependence, at $\rho^2 \simeq 2\%$ of the variance, which is the same magnitude recovered here. The picture is consistent across four model classes and two independent methodologies: about a tenth of the variance in γ is accounted for by the clustering amplitude of the input field, $\sim 14\%$ by the spectrum as a whole, a further $\sim 2\% - 3\%$ by the astrophysical parameters, and no method recovers more. The ceiling is therefore set by the information content of the inputs rather than by the flexibility of the regressor.

Why γ is only partially predictable

The ceiling encountered above has a partial explanation in the structure of the loss, and the derivation of Appendix E provides it. There, the coefficient $b(\boldsymbol{k}, \boldsymbol{q})$ of Eq. (E.1) is defined by dividing the three-point average appearing in that term by $\sqrt{P_{\rho^\gamma}(q)}$, the typical amplitude of the source modes, so that $\tilde{G}(q)\sqrt{P_{\rho^\gamma}(q)} = J(q)$ can be factored out; however, this division removes the scaling of the average with the magnitude of those modes, but leaves behind a normalized cross correlation, still dependent on their phases. Since raising a field to a power is a non-linear pointwise operation, changing γ alters both the amplitudes of the source modes, which $P_{\rho^\gamma}(q)$ captures and $J(q)$ therefore absorbs, and their phases, which it does not. In this sense $J(k)$ fixes the amplitude of the correction as a function of scale, while γ additionally controls, through those phases, how well the applied displacement is aligned with the discrepancy it is meant to remove. This residual

dependence on γ explains why it is recovered reproducibly across initializations at fixed target, whereas α, k_l and ν are constrained only through the combination that forms $J(k)$. This suggests that the regressions of this section, among other possible effects, reach the ceiling they do due to a limitation of the summary supplied to them. The input power spectrum alone does not carry the phase information needed for the spatial alignment between the source field and the residual present in the coefficient $b(\boldsymbol{k}, \boldsymbol{q})$: a regressor given the $P_\rho(k)$ of the input field can therefore access only the phase-independent part of what determines γ. This also accounts for the insensitivity to $k_{\max}$ noted above, since extending the cutoff supplies more modes but no phase information. A regressor operating on the full input field ρ rather than on its spectrum, e.g. a convolutional network on ρ, would have access to the phase information that the summaries used here discard. This was not attempted here. Nonetheless, one contribution is genuinely absent from the input: the galactic wind energy $\mathrm{A_{SN1}}$ modifies the reference, and hence the residual, while leaving the FastPM field untouched, thus no functional of it can encode feedback, which is why the corresponding component of $\boldsymbol{\phi}$ survives conditioning on $\hat{\sigma}_2$ and contributes the small dependence found in Section 3.4.2.

A final observation relates to how much of the scatter in γ is identified at all. Repeating the optimizations on 6 of the CV simulations with γ held fixed at 0.75, close to the typical recovered value, yields, in every case, best achieved losses greater than those of the free-γ case, with $\sim 0\% - 13\%$ differences across the realizations tested. Fixing γ therefore does carry a cost, but a small one against the spread in best achieved loss between realizations. This suggests that the scatter of γ about a typical value may actually carry little information about how well the model can perform; however, with 6 realizations and a comparison confined to a single fixed value this is a hypothesis rather than a result, and a proper test would require imposing γ at several values across a larger set of realizations. A related conclusion is reached from a different direction in Section 3.6.2, where the performance of a transferred parameter set is found to be rather insensitive to γ over the range in which the recovered values cluster.

Four distinct contributions to the scatter in γ can thus be identified:

1. the clustering amplitude and scale dependence of the input field, accounting for $\sim$15% of the variance, established independently in Section 3.4.2 and by the regressions of this section, and anticipated by the hypothesis of Section 3.4.1;

2. the astrophysical parameters, and in particular $\mathrm{A_{SN1}}$, contributing a further $\sim$2%–3% (Section 3.4.2);

3. a component that is not reproducible across independent optimizations of the same target, and thus not associated with the input field. Its presence is visible in Figure 3.17, where γ retains some scatter across the optimizations of a fixed realization.

In principle, its contribution could be measured as the ratio between the variance of γ across different optimizations on a fixed simulation and the total variance across the suite, but the present data is not sufficient to determine it, as it uses only 3 seeds per simulation, some of which may lead to failed optimizations;

4. information about the input field which the summaries employed here lack. The mechanism expected to carry it is the spatial alignment between the source term and the residual, which depends on the phases of the input field and is therefore invisible to $P_\rho(k)$, as the argument of Appendix E anticipates. This contribution is also not quantified.

Together, the last two account for the remaining $\sim 80\%$ of the variance in proportions this analysis cannot separate. Nonetheless, it is important to reiterate that the volume of the simulations is also relevant here. In a $25\,h^{-1}$Mpc box, the densest regions, on which ρ^γ is concentrated (especially for larger γ), comprise a small number of objects whose arrangement varies considerably between realizations; the alignment between the source term and the residual is therefore strongly realization-specific, and it is exactly this contribution that a spectrum-based summary discards. Consistently with this reading, the tendency of γ to cluster near values of the form $i/4$, noted in Section 3.3.3, is most naturally attributed to the optimization rather than to the fields, since no corresponding feature appears in any statistic of the input.

3.5.2 Predicting $J(k)$

Contrarily to γ, for $J(k)$ the cosmological and astrophysical parameters $\boldsymbol{\phi}$ contribute significantly. Comparing the performance of a neural network across ten splits between $P_\rho(k)$-only and ($P_\rho(k)$, $\boldsymbol{\phi}$) inputs, the two choices are barely distinguishable, with a mean R^2 across the splits of 27.5% including $\boldsymbol{\phi}$ against 18.1% without, but the standard deviation between splits being $\sim 9\%$ in each case, as large as the difference itself. Compared on identical splits, however, the mean improvement is 9.4% with a standard error of 0.8%, and it is positive in all ten repeats. Ridge regression on the same splits gives a median gain of $\sim 8\%$, positive in every one of thirty splits. Two model classes with quite different inductive biases therefore agree on both the sign and the size of the effect, and neither produces a single split in which the additional inputs are harmful. The contrast between the marginal and the paired comparison is the reason the paired design was adopted throughout: at this sample size an effect of ten percentage points is invisible in the marginal distributions but glaring in the differences.

Therefore, $J(k)$ encodes appreciably more information about the cosmological and astrophysical parameters than γ does. Since $J(k) = \tilde{G}(k)\sqrt{P_{\rho^\gamma}(k)}$ and the γ-dependence is confined to the second factor, this seems to indicate that this additional predictability

resides mostly in the Green's function itself, i.e. $\tilde{G}(k)$ mostly relates to $\boldsymbol{\phi}$ in a way a regressor can exploit, rather than the exponent of the source term. This has not been verified directly, as doing so would require regressing $\tilde{G}$ itself, which is precisely the badly-conditioned quantity that motivated introducing $J(k)$ in the first place, and is therefore presented as an interpretation rather than a result.

Why $J(k)$ is only partially predictable

Figure 3.25 suggests where the difficulty in regressing $J(k)$ may originate: the spread of this function across the CV set, at fixed cosmology, overlaps substantially with that across the LH set, at varying cosmology, thus the signal a regressor must extract is of the same order as the realization noise carried by each individual training example. This limitation is tightly related to the experimental design rather than to the method: the LH set provides neither repeated draws at fixed $\boldsymbol{\phi}$, nor a set of simulations with varying cosmology at fixed seed, and so the two contributions cannot be separated; optimizing more simulations of the same kind would reduce the variance of the estimator without improving the ratio of signal to realization noise in any single example, which is what limits the inference here. Recovering the dependence more completely would require changing those conditions, i.e. either averaging over several realizations at each point in parameter space, fixing the seed of the initial conditions across cosmologies, or working with larger volumes; Section 3.8 describes what datasets would be strongly beneficial for this analysis and what is currently available.

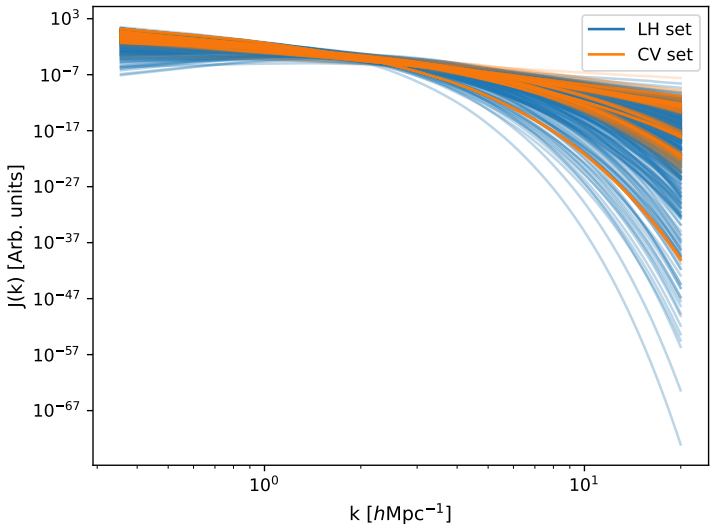


Figure 3.25: $J(k)$ for the optimizations on the LH set introduced in Section 3.3.4 (blue curves) and for the ones on the CV dataset of Section 3.3.3 (orange curves).

3.5.3 Limitations

It is important to highlight some limitations of the analysis above. First, the values quoted for the neural networks are minima of the validation loss over training and are therefore mildly optimistic as estimates of generalization error; nonetheless, ridge and Gaussian process regression, which involve no such selection, reproduce them for both targets, indicating that this bias is small in any case.

Additionally, since $J(k) = \tilde{G}(k)\sqrt{P_{\rho^\gamma}(k)}$ and the source power spectrum is itself a functional of the input field, part of the $J(k)$ target is reconstructible from the input given γ, so a regressor can recover a fraction of the variance in $J(k)$ without having learned anything about the model. For this reason, the absolute values for $J(k)$ should be read with caution; on the contrary, the paired comparison is immune to this effect, as the reconstructible part is common and thus cancels in the difference. The absolute figures then overstate how much is being learned about $\tilde{G}(k)$. Moreover, the loss on $J(k)$ is a sum over wavenumber bins, of which there are considerably more at large k, so that increasing $k_{\max}$ reweights the objective towards small scales as well as supplying additional information; the comparisons between inputs, made at fixed $k_{\max}$, are unaffected.

The composition of these two halves of the emulation scheme was not tested. The best results achieved in terms of prediction of γ and $J(k)$ only account for $\sim 15\%$ of the variance for the former and $\sim 28\%$ for the latter; since the two predictions enter $\tilde{G}(k) = J(k)/\sqrt{P_{\rho^\gamma}(k)}$ in combination, their errors compound, and the accuracy of the fields produced by a model built from predicted rather than fitted quantities has not been measured. This is the natural next step, and is returned to in Section 3.8.

Finally, the configuration predicting the full $\boldsymbol{\theta}$ vector with a loss defined on $J'(k < k_{\max})$ never outperformed the sample mean. This may be due to two reasons: first, $J'(k)$ is a numerical approximation to $J(k)$, so that its use as a target adds the error of Eq. (3.4) on top of a quantity already only weakly predictable; second, placing the loss on $J'(k)$ removes the degeneracy from the metric but not from the output space: the network still has to predict α, k_l, ν and γ, whose individual values are determined only up to the transformations characterized in Section 3.3.3, and therefore the objective retains flat directions in the quantities actually being predicted.

3.6 Performance at fixed parameters

The regressions discussed in the previous section can explain only a small part of the variance of γ and $J(k)$. Establishing a mapping $\boldsymbol{\phi} \mapsto \boldsymbol{\theta}$ (or rather $\boldsymbol{\phi}$, $P_\rho(k) \mapsto \gamma$, $\tilde{G}(k)$) accurate enough to be used in place of an optimization therefore remains out of reach. For this reason, it is worth asking the complementary question, i.e. how much is actually lost by not having such mapping and applying `JERALD` with a single fixed parameter set to

every simulation. At fixed cosmology, parameters optimized on one CV realization were applied to the remaining ones and found to generalize without appreciable degradation, consistent with the invariance of $J(k)$ under cosmic variance documented in Section 3.3.3 (see, for instance, Figure 3.16). The real test, however, considers varying cosmology. For this purpose, the parameters optimized on a single LH simulation (LH760[4]) were thus applied, unchanged, to 100 additional LH simulations spanning the full prior range in Ω_m, σ_8 and the four astrophysical parameters, none of which was seen during optimization. LH760 was not chosen at random: it was selected, from among the 500 optimized parameter sets, as the one generalizing best over another 100 unseen simulations (different with respect to the 100 included in the results below) according to the k-averaged fractional residual power $\Delta^2 = \langle \Delta^2(k) \rangle$, where

$$\begin{aligned} \Delta^2(k) &\equiv \frac{\left\langle \left| \delta_{\rm pred}(\boldsymbol{k}) - \delta_{\rm targ}(\boldsymbol{k}) \right|^2 \right\rangle_{|\boldsymbol{k}|=k}}{\left\langle \left| \delta_{\rm targ}(\boldsymbol{k}) \right|^2 \right\rangle_{|\boldsymbol{k}|=k}} \\ &= 1 + T^2(k) - 2\, r(k)\, T(k) \quad , \qquad T(k) \equiv \sqrt{\frac{P_{\rm pred}(k)}{P_{\rm targ}(k)}} \quad , \end{aligned} \tag{3.11}$$

where $r(k)$ is the cross correlation coefficient. The resulting power spectra are shown in Figure 3.26 and compared, in the right panels of the same figure, against the 500 optimizations for which `JERALD` was instead individually optimized (referred to as "train" below).

These plots highlight two important features. First, the improvement over the input simulation is significant and consistent: with fixed parameters, `JERALD` holds the power spectrum to within 10% of the reference out to $\sim 7\,h\mathrm{Mpc}^{-1}$, whereas plain FastPM has already departed by $\sim 50\%$ at $\sim 4\,h\mathrm{Mpc}^{-1}$, and the cross correlation coefficient crosses 0.9 at $\sim 7\,h\mathrm{Mpc}^{-1}$ rather than $\sim 4\,h\mathrm{Mpc}^{-1}$. Second, the fixed-parameter results are statistically comparable to those obtained with per-simulation optimization: the mean of the test set lies within the scatter of the train set across essentially the whole range of scales.

The same comparison for the bispectrum is shown in Figures 3.27, for configurations with arm lengths $k_1 = k_2$ and fixed angle $\pi/3$ (the cases $\pi/2$ and $\pi/4$ were also tested, giving similar results). Importantly, the model is optimized against a voxel-space L_2 loss, with virtually no guarantee of reproducing higher-order statistics (as was observed in Section 3.1.2, where improved losses coexisted with degraded power spectra and bispectra); Appendix E does establish a correspondence between the real-space loss and Fourier-space quantities, but this does not by itself constrain higher-order statistics: the

[4] For reference, the best model parameters for LH760 are $\boldsymbol{\theta}_{\rm LH760} = (10.15, 0.4356, 2.5, 0.9171)$ and the corresponding $\boldsymbol{\phi}_{\rm LH760} = (0.3682, 0.6734, 1.09429, 0.66619, 1.14949, 0.87965)$; the ranges for these parameters can be found here.

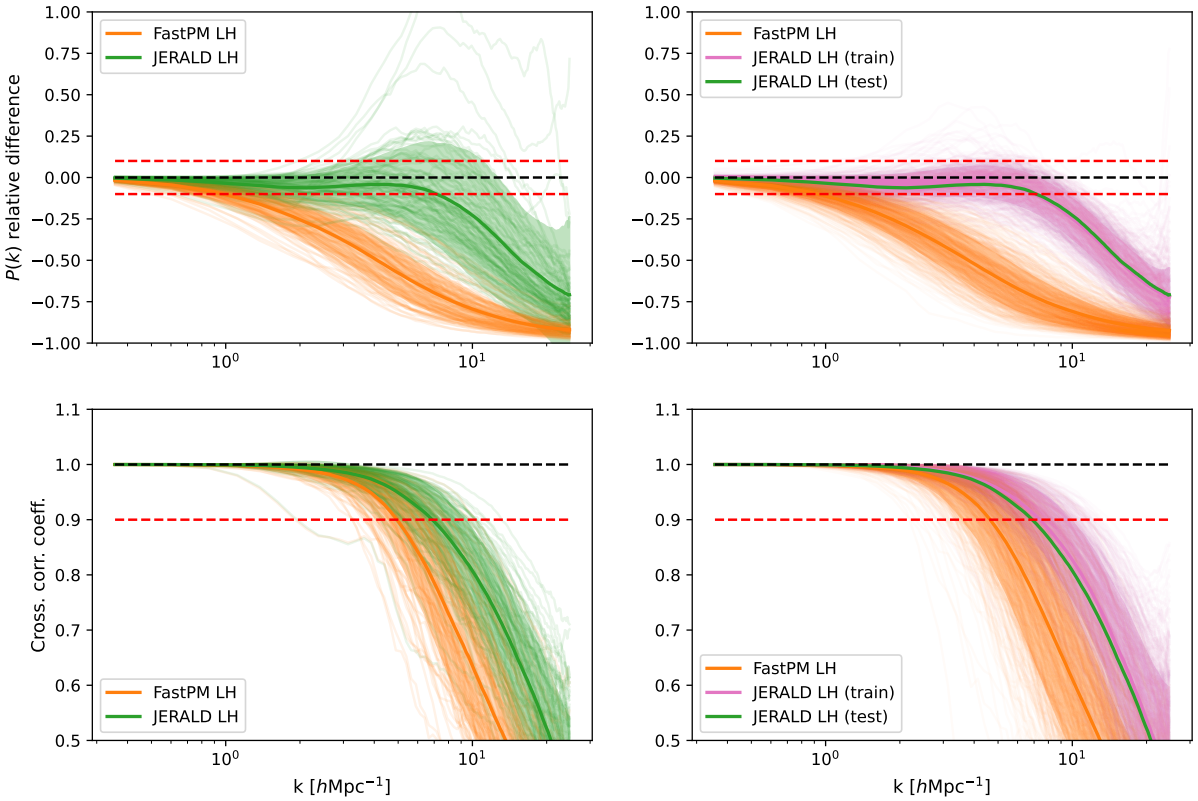


Figure 3.26: Performance of `JERALD` with a single fixed parameter set on 100 simulations it was not optimized on. *Top row*: relative difference between the power spectrum of the output and that of the reference. *Bottom row*: cross-correlation coefficient between output and reference fields. *Left column*: the 100 test simulations of the LH set, comparing the uncorrected FastPM input (orange) with `JERALD` (green). *Right column*: same as above, with the addition of the 500-simulations train set (magenta), in which each simulation carries its own individually optimized parameters. Lines in transparency show individual simulations, thick lines their mean, and shaded areas represent $\pm 1\sigma$ regions.

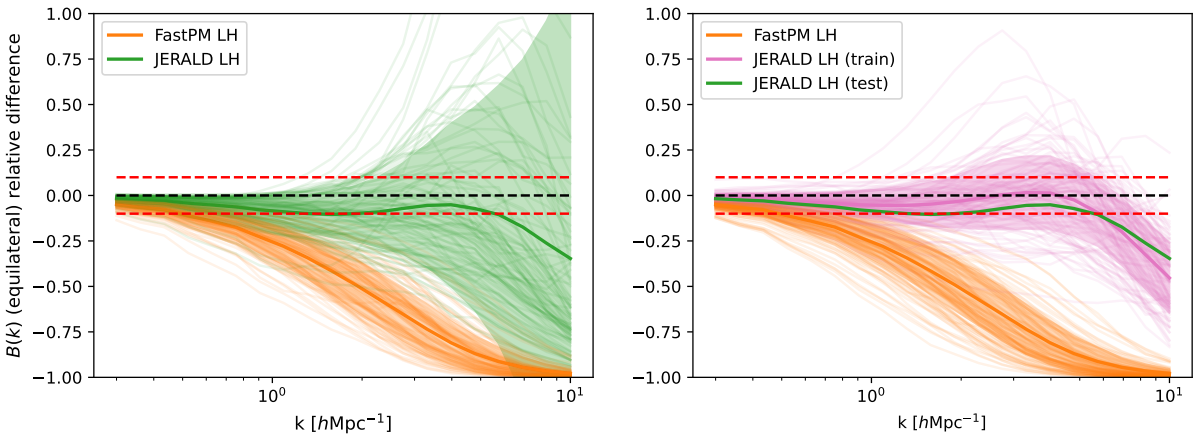


Figure 3.27: Same as Figure 3.26 for equilateral triangle bispectra (equal arm lengths, $\theta = \pi/3$).

loss is quadratic in the fields, and while the cross term appearing in that derivation is a three-point object correlating two modes of the input source field with one mode of the residual, the bispectrum of the output itself is not constrained. Agreement at the level of $B(k)$ is therefore not guaranteed by construction and has to be checked.

On average, the bispectrum relative difference remains within 10% of zero out to $\sim 6\,h\mathrm{Mpc}^{-1}$, while FastPM loses $\sim 90\%$ agreement over the same range: the fixed-parameter generalization is therefore not an artifact of restricting attention to two-point statistics. The realization-to-realization scatter, however, is considerably larger than for the power spectrum. This is expected, as the bispectrum is intrinsically harder to estimate in these volumes than the power spectrum, as far fewer triangle configurations are available at a given scale (especially for longer arm lengths) in a $25h^{-1}$Mpc box. Additionally, and most importantly, as mentioned above, the model is not optimized against any three-point statistic, so there is no mechanism by which the optimization would control its variance. Note, however, that that the relative difference $(B_{\texttt{JERALD}} - B_{\mathrm{CAMELS}})/B_{\mathrm{CAMELS}}$ is bounded below by -1 but unbounded above: the FastPM band, pinned near -0.9, is therefore compressed by construction, while `JERALD`'s, sitting near zero, is not. Some of the visual contrast between the two bands is therefore an artifact of the choice of the y axis.

In conclusion, the model evaluated with a single fixed parameter set brings a substantial improvement over the uncorrected input, while the further improvement brought by optimizing $\boldsymbol{\theta}$ on each simulation individually is small in comparison; this, however, holds only at the level of the mean, and for two-point statistics. A single well-chosen parameter set reproduces the reference power spectrum to within 10% over a wide range of scales on simulations it was never fitted to, which is sufficient for applications operating at that level, but not further: the corresponding bispectra are recovered with little bias but with considerably larger realization-to-realization scatter, so a universal parameter set is adequate for the mean three-point signal and much less so for any individual realization of it. The same distinction matters for field-level applications such as those of Chapter 4, where what is required is the accuracy of each map rather than of an average over many. For those purposes a genuine dependence of $\boldsymbol{\theta}$ on $\boldsymbol{\phi}$ would still be valuable; the results of this section should accordingly be taken as a simplification only for specific purposes, while the fact that this dependence could not be recovered in Section 3.5 remains a limitation. Independently of this, the results in this section point to a procedure that was not attempted here: rather than optimizing each simulation separately and then selecting or interpolating among the recovered values, one may optimize a single parameter set jointly against an ensemble of simulations, minimizing the loss summed over a batch of targets spanning the prior range, thus averaging over realization noise instead of inheriting it. This may also improve the regression problem itself: a jointly optimized $\boldsymbol{\theta}^\star$ absorbs the large component common to all simulations, so that the residuals $\boldsymbol{\theta}_i - \boldsymbol{\theta}^\star$ (or rather the

corresponding residuals in $J(k)$) contain only the putative dependence on $\boldsymbol{\phi}$ together with realization noise, making it a better-posed regression target than $\boldsymbol{\theta}_i$ itself.

3.6.1 Nearest neighbor interpolation

As the previous section shows, a well-chosen fixed parameter set generalizes across cosmologies; however, a typical one does not, as this section reveals. Figure 3.28 shows the result of using nearest neighbors as interpolation strategy, in which each test simulation of the previous figures is assigned the parameters $\boldsymbol{\theta}$ of its nearest neighbor in the $\boldsymbol{\phi}$ space among the 500 train simulations[5]. For completeness the same test was repeated using only (Ω_m, σ_8), leading to the same conclusions, so that what follows does not depend on this choice.

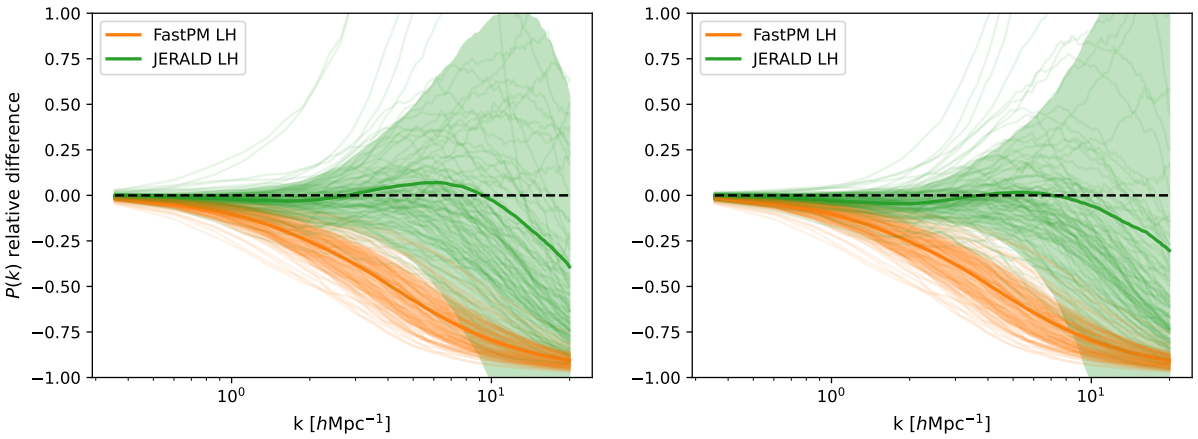


Figure 3.28: Performance of `JERALD`, in terms of power spectrum relative difference with respect to the reference CAMELS simulation, on 100 simulations it was not optimized on, where each simulations inherits the model parameters of the closest one in the $\boldsymbol{\phi}$ parameter space. *Left*: the distance is computed on the entire $\boldsymbol{\phi}$ vector; *right*: the distance is computed only using the input (Ω_m, σ_8). Similarly to Figures 3.26 and 3.27, lines in transparency show individual simulations, thick lines their mean, and shaded areas represent $\pm 1\sigma$ regions.

The mean behavior remains reasonable, but the scatter is very large, with individual realizations departing by more than 100% at intermediate scales. Since the LH dataset samples a six-dimensional space with 1000 points and only 500 were used as reference, the nearest neighbor of a given test point may be far from it; however, more importantly, each $\boldsymbol{\theta}$ in the 500 reference set is a single draw from the degenerate basin characterized in Section 3.3.3, therefore selecting one by cosmological proximity inherits that scatter.

[5] Note that, as mentioned in Section 3.5, the six components of $\boldsymbol{\phi}$ are sampled over ranges differing by up to a factor of nine, and the four feedback parameters are sampled log-uniformly rather than uniformly; for this reason, distances were computed after normalizing (or log-normalizing) each component to the range $[-1, 1]$.

Note that this experiment uses the single nearest neighbor, rather than an average over the k nearest as would be more customary. The reason behind this is that averaging is not a well-defined operation in the $\boldsymbol{\theta}$ parameter space: Section 3.3.3 show that the optimized parameters are confined to curved degeneracy lines along which the model's predictions are essentially unchanged, and a convex combination of two points on such a locus does not, in general, lie on it, thus the arithmetic mean of two equivalent parameter sets may perform worse than both. The same conclusion follows from the functional form alone, since, e.g., k_l enters $\tilde{G}(k)$ through $1/k_l$. The quantity it would be meaningful to average is $J(k)$ itself; however, averaging $J(k)$ over neighbors and then recovering the parameters that reproduce it is practically the regression attempted in Section 3.5.

3.6.2 The role of γ and $J(k)$ in transferability

Figure 3.29 shows the generalization metric Δ^2 (see Eq. (3.11)) evaluated on the same 100 unseen simulations used above to find LH760, against the value of γ optimized for each of the 500 train simulations.

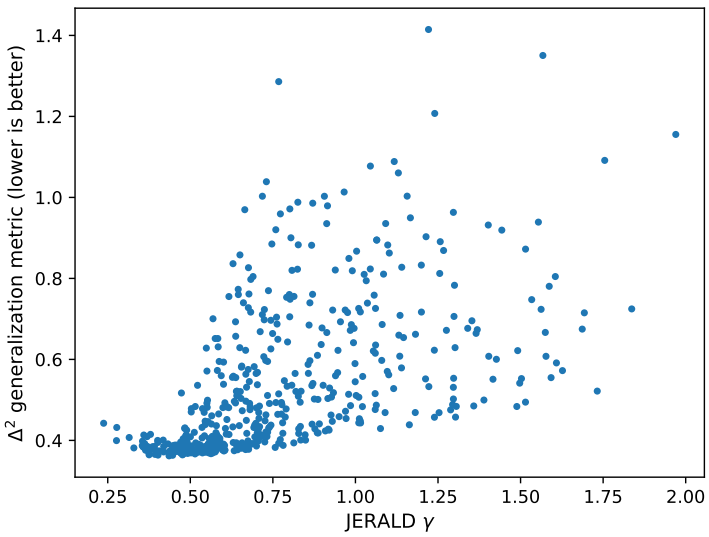


Figure 3.29: Generalization performance of the 500 LH train simulations against the recovered value of γ. The vertical axis is the fractional residual power of Eq. (3.11) over a common set of 100 unseen simulations, lower being better. Note that, since every parameter set is scored on the same 100 simulations, differences between points reflect the fitted parameters alone and cannot be attributed to some sets having been evaluated on more difficult boxes.

The dependence is practically one-sided: around $\gamma \sim 0.5$ the residual power sits on a floor, while above this value both the floor and the scatter rise, the worst sets exceeding three times the floor value. A large γ is therefore not merely worse on average but also less predictable in how it transfers, as might be expected of a source term ρ^γ increasingly

dominated by the extreme tail of the density field, and hence by the particular realization it was fitted to.

It is worth asking whether this dependence acts through $J(k)$ rather than through γ directly. Conditioning on a scalar summary of each set's $J(k)$, taken as its logarithmic departure from the population median over $k < 5\,h\,\mathrm{Mpc}^{-1}$, leaves the correlation between γ and Δ^2 essentially unchanged, at $\rho_S = 0.71$ marginally against $\sim$0.6 conditioned; the reverse conditioning leaves the summary with almost no dependence of its own beyond a weak residual. γ therefore accounts for most of the variation in transferability, which is what the analysis of Appendix E would lead one to expect: $J(k)$ absorbs the dependence of the loss on the magnitude of the source modes, whereas γ retains information on their phases, and hence on the spatial alignment between the applied displacement and the discrepancy it is meant to remove, and alignment with the structures of one particular realization is precisely what does not carry over to another. Nonetheless, the comparison is indicative rather than rigorous, both because a summary of $J(k)$ naturally carries less information about the function and because what is transported to a new simulation is $\tilde{G}(k)$ rather than $J(k)$.

The values recovered by the optimizations concentrate mostly between roughly 0.4 and 0.7, i.e. largely within the region where the residual power is at its floor. Over this flat region, there is little to be gained by replacing a population value of γ by the value fitted to an individual simulation: the conclusion relevant to the emulation scheme is therefore that γ may not need to be predicted at all: holding it at a typical value appears to cost little, at least for simulations of the size and resolution considered here, consistently with the fixed-γ comparison on the CV set of Section 3.5. If so, the quantity left to emulate is $J(k)$ alone; this was not tested.

3.7 Performance limits

The aim of this section is to identify what determines how well the model performs on a given simulation. Figure 3.30 shows the lowest loss attained against the realized clustering amplitude of the input field, $\hat{\sigma}_2$, for both the 500 LH simulations introduced in Section 3.3.4 and 11 CV simulations of Section 3.3.3. The correlation is strong: simulations containing more non-linear structure are systematically harder to fit, with the attained loss varying by more than an order of magnitude across the range. The CV and LH points lie on the same relation: whether the amplitude of a given box is large because of its cosmology or because of chance is unimportant, and the loss depends on the realized amplitude itself.

The same conclusion holds at the level of the output fields. Figure 3.31 shows the cross-correlation coefficients for the LH set before and after applying `JERALD`, colored

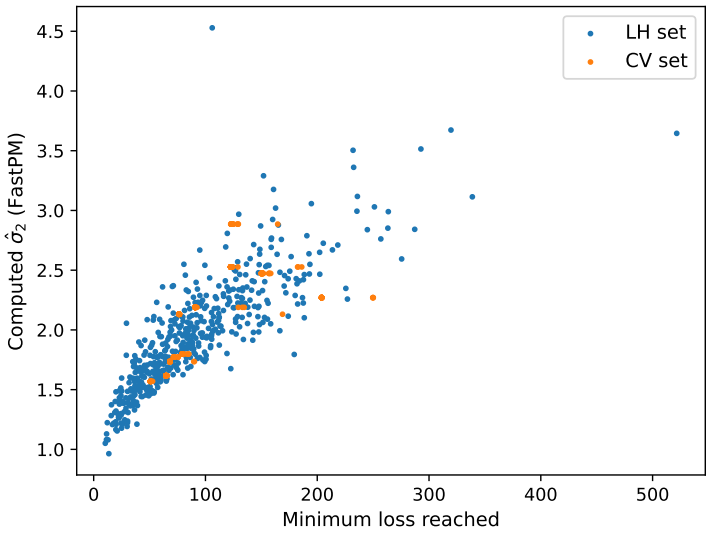


Figure 3.30: Scatter between minimum loss reached in `JERALD` optimizations and the realized input FastPM amplitude $\hat{\sigma}_2$ for 500 LH simulations (see Section 3.3.4) and 11 reference CV simulations (see Section 3.3.3).

by $\hat{\sigma}_2$. The curves are ordered by it in both panels: the model raises every one of them without removing the ordering, as expected if the clustering amplitude governs the degree of non-linearity at a given scale, and thus both the failure of the approximate solver and the fraction of it that an effective displacement model can recover.

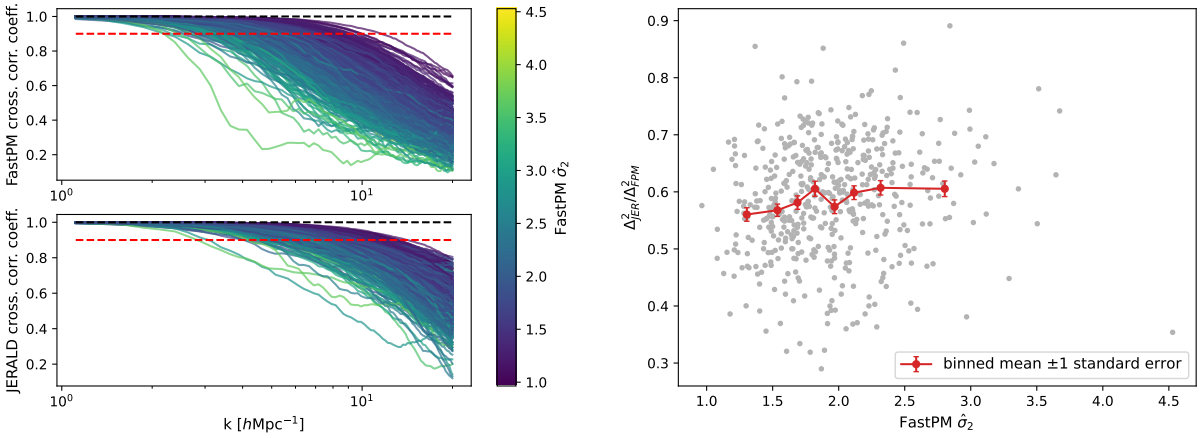


Figure 3.31: *Left:* cross correlation coefficients between FastPM (top) and `JERALD` (bottom) and 500 LH simulations (see Section 3.3.4), colored by the realized FastPM amplitude $\hat{\sigma}_2$. *Right:* scatter between the performance improvement of `JERALD` over its FastPM input quantified via the metric in Eq. (3.11) and the realized FastPM amplitude $\hat{\sigma}_2$ for 500 LH simulations (see Section 3.3.4).

The improvement is close to uniform across the sample: the ratio $\Delta^2_{\rm JER}/\Delta^2_{\rm FPM}$, where Δ^2 is the metric defined in Eq. (3.11) (averaged over k), also included in Figure 3.31, sits at $\simeq 0.6$ with no appreciable trend in $\hat{\sigma}_2$, so that, on average, the residual field power is reduced by roughly 40% irrespective of how non-linear the box is. This is the quantitative

form of the ordering noted above: the model does not preferentially rescue the boxes in which the approximate solver fails worst, but improves all of them by a comparable factor.

3.8 Summary and outlook

The purpose of this chapter was to test whether `JERALD` could be promoted from a model fit to one single fixed cosmology, as in Chapter 2, into a full cosmology-dependent emulator. Two routes were pursued: conditioning the architecture directly on $\boldsymbol{\phi}$ (Section 3.1), and identifying a functional relation $\boldsymbol{\theta}(\boldsymbol{\phi})$ that could be evaluated rather than re-fitted (Sections 3.2 through 3.7). Neither has so far produced a working cosmology-conditioned model. Nonetheless, what the analysis in this chapter produced is an account of the model's behavior, revealing what its loss does and does not determine as well as confirming connections, predicted by physical arguments, between its parameters and properties of the simulations it uses as inputs and outputs.

Concerning the first route, the direct integration in the model of a neural network outputting $\boldsymbol{\theta}$, so that its weights are trained rather than $\boldsymbol{\theta}$, failed due the physical constraints the parameters must satisfy and the cost of training against many targets simultaneously (Section 3.1.1). Replacing the Green's function with a spline revealed an important property of the functional form of Eq. (2.7): across every variant tested, added flexibility improved the voxel-space loss while degrading the power spectrum of the output field (Section 3.1.2); trading flexibility for physical interpretability then appears to be the strength of Eq. (2.7), rather than it being a convenient parametrization among many. The comparison with [91] supports this reading: there the spline kernel enters as a correction, so that the physical solution is recovered when it vanishes, and the objective includes a power spectrum term. Both remedies, only briefly tested here, suggest promising avenues for pursuing this approach.

The second route requires $\boldsymbol{\theta}$ to be well determined at fixed $\boldsymbol{\phi}$; this analysis shows that this is not the case. Varying redshift alone produced multiple competing minima, directions along which the loss is nearly flat, and an optimum that relocates discontinuously rather than tracking a smooth trajectory (Section 3.2.2). Across the CAMELS cosmic variance set the same structure appears in a sharper form: at fixed cosmology and fixed initial conditions, independent optimizations recover widely scattered values of α, k_l and ν that describe almost indistinguishable corrections (Section 3.3.3). Following this degeneracy to its origin is what produced the chapter's main positive result. The combination

$$J(k) = \tilde{G}(k)\sqrt{P_{\rho^\gamma}(k)}$$

is reproducible where its constituents are not, and Appendix E highlights the reason: since $\tilde{G}(k)$ depends on the modulus of the wavevector alone, $J(k)$ is the unique combination

for which the remaining factor carries no amplitude information, so the loss can depend on the model only through $J(k)$ and through the phases of the source field. The strongly degenerate fit of the previous sections thus turns into one over a single low-dimensional function of scale. The same analysis also explains the limits encountered in predicting the remaining parameter γ and $J(k)$. For the former, a potential physical connection between its value and properties of the input field put forward in Sections 3.4.2 is confirmed by Appendix E and verified concretely in Section 3.5, where the clustering amplitude of the input field accounts for $\sim 10\%$ of its variance and the full input power spectrum for $\sim 15\%$, while including the astrophysical parameters supplies an additional $\sim 3\%$. The ceiling is set by the information content of the inputs rather than by the flexibility of the regressor, and Appendix E identifies what is missing: the residual dependence on γ is carried by the spatial alignment between the source term and the residual, which depends on phases and, as such, is invisible to any summary built from the input $P_\rho(k)$. For $J(k)$, the cosmological and astrophysical parameters contribute significantly to its predictability, by $\sim 10\%$ of explained variance in paired comparisons, suggesting that $\tilde{G}(k)$, rather than γ, absorbs much of the model dependence on $\boldsymbol{\phi}$; however, the spread of $J(k)$ at fixed cosmology is comparable to its spread across the prior range of cosmologies, so that the signal to be extracted sits on realization noise of its own size in every training example. This limitation is associated to the experimental design rather than to the method.

A practical fallback is thus investigated: a single well-chosen $\boldsymbol{\theta}$ trained on a set of $\boldsymbol{\phi}$ and applied unchanged to unseen simulations at different $\boldsymbol{\phi}$ yields power spectra within 10% out to $k \simeq 7\,h\mathrm{Mpc}^{-1}$ and (equilateral) bispectra in the same range out to $k \simeq 6\,h\mathrm{Mpc}^{-1}$ (Section 3.6). Nonetheless, this holds at the level of the mean, while the realization-to-realization scatter is considerably larger. A fixed parameter set is therefore sufficient for applications operating on ensemble averages and two-point statistics, but inadequate for those requiring the accuracy of individual maps, e.g. the regime of Chapter 4.

The analysis in this chapter then suggests several additional experiments. The most beneficial improvement, however, is a change of data rather than of method. The results presented above would considerably benefit from the availability of two datasets, varying the same parameters as LH but in larger volume, in order to remove many of the small-box-size effects discussed throughout the chapter: one in which the seed of the initial conditions is held fixed across cosmologies, so that the mapping between $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ could be examined without realization-to-realization noise superimposed on it, and one with multiple realization per $\boldsymbol{\phi}$, in order to quantify how much this noise affects the determination of the dependence. The CAMELS suite does provide a dataset (the SB, or "Sobol sequence" set) similar to LH with boxes of size $50\,h^{-1}\mathrm{Mpc}$[6], although varying 30 astro-

[6] This dataset was released only few months prior at the time this thesis was written and could not be exploited.

physical parameters rather than 4, with roughly the same number of simulations (1024), so that the sampling per dimension is considerably sparser than in the LH set used here.

Yet, the results above still suggest further information can be extracted from the data available at the time of writing. For what concern the determination of γ, the ceiling on predicting it is attributed here to phase information that $P_\rho(k)$ discards. This is directly testable: a regressor operating on the input field ρ itself, e.g. a convolutional network, would have access to exactly the information the summaries used here discard. The same experiment would bound the irreproducible component of the scatter, which the three optimizations per simulation on the LH set performed here are too few to isolate. Conversely, performing more optimizations per $\boldsymbol{\phi}$ would allow to estimate the contribution of this component and therefore determine the amount of information on γ contained in the phases of the input field.

In an opposite direction, γ may not need to be predicted at all. Transferred performance is rather flat in γ over the range in which the fitted values cluster, degrading only above it (Section 3.6.2), thus fixing γ at a typical value costs little in transferability. Emulating $J(k)$ alone at fixed γ would then reduce the problem to its better-determined half, the same half in which $\boldsymbol{\phi}$ demonstrably contributes. Running further optimizations on the LH set at different fixed γ values would allow to test this claim.

Moreover, the composition of the two halves of the emulation scheme was never evaluated: predictions of γ and $J(k)$ enter $\tilde{G}(k) = J(k)/\sqrt{P_{\rho^\gamma}(k)}$ in combination, so their errors compound, and the accuracy of fields produced from predicted rather than fitted quantities has not been measured. That measurement is the natural next step.

Finally, these results motivate the choice made in the next chapter. Since a cosmology-conditioned `JERALD` is not available, the simulation-based inference pipeline of Chapter 4 is built instead around a neural emulator conditioned on $\boldsymbol{\phi}$ by construction. How the two lines of work might be combined is discussed in Chapter 5.

Chapter 4

Simulation Based Inference on fields

The work presented in this chapter is the outcome of a collaboration, published in [2], of which the author of this thesis is third author. The inference pipeline, the neural architectures and the statistical analysis were developed principally by the first two authors; the specific contribution of the present author consisted in producing the approximate N-body simulations on which the pipeline operates, i.e. the FastPM runs described in Section 4.2.2, and in the continued discussion of the design, execution and interpretation of the analysis with the first two authors over the course of the project. The chapter addresses, using an existing conditioned emulator in place of `JERALD`, precisely the field-level inference problem that motivated the work of the preceding chapters; Chapter 5 returns to how the two lines of work might eventually be combined. Only the parts of [2] relevant the subject of this thesis are reproduced here: part of the introductory material has accordingly been omitted as it duplicates Chapter 1. Additionally, Section 4.4, an appendix in [2], has here been promoted to a section, as it discusses results relevant for this thesis, which Chapter 5 will return to.

4.1 Introduction

The computational cost of full-hydrodynamical simulations mentioned in Chapter 1 poses a fundamental obstacle for traditional inference pipelines, in which Markov Chain Monte Carlo approaches [192, 193] require the evaluation of a likelihood function at every sampled point in the high-dimensional parameter space relevant for modern cosmology. The standard solution is to adopt approximate likelihoods, typically assumed to be Gaussian, defined in terms of a summary statistics, which are often limited to the power spectrum. While computationally tractable [194], these approaches are known to discard a significant fraction of the cosmological information encoded in the non-Gaussian features of the matter and tracer fields, especially on small and intermediate scales. For example, [195] shows that, on a simplified analytic model based on lognormal fields, often used as an ap-

proximation to the late-time cosmological density field, standard approximations adopted to model a likelihood for correlation functions can cause significant biases, whereas field-level inference gives more accurate and precise posteriors, with gains increasing with non-Gaussianity. Similar conclusions have been reached by [196] in regards of cosmic shear data. However, the exact magnitude of the gain of full-field analyses with respect to Gaussian summary-statistics likelihoods is context-dependent: crucial factors include the degree of non-Gaussianity and the scale considered [197, 195], the characteristics of the instrument adopted and whether the summary statistic is actually close to sufficient (see e.g., [198, 199]).

In the last few years, Simulation-Based Inference (SBI) has emerged as a powerful framework that circumvents the need for an explicit likelihood function (see e.g. [200, 201] for reviews). Instead of specifying a likelihood, SBI relies on forward simulations and machine learning methods to learn the posterior distribution, the likelihood function or the likelihood-to-evidence ratio of parameters θ given the data x. This makes SBI particularly well-suited for cosmology, where numerical simulations are already central to modeling the non-linear regime of structure formation [202], but where analytical likelihoods are often intractable. Early applications to weak-lensing maps showed that likelihood-free inference can combine forward modeling, realistic survey masks and noise, and neural compression of summary statistics [203].

SBI has already been demonstrated to be a powerful approach to inference in a range of cosmological and astrophysical applications. For example, [204] used neural classification for Bayesian model comparison in Type Ia supernova analyses, demonstrating that SBI can marginalize over thousands of latent variables. In cosmological field reconstruction, [205] introduced an SBI-based Bayesian framework applicable to generic non-differentiable forward simulators, enabling sampling from the posterior of the underlying field. In 21-cm cosmology, [206] applied truncated marginal neural ratio estimation to mock observations, showing that SBI can effectively handle complex foreground, beam, and noise effects while targeting marginal posteriors efficiently. In studies of galactic binaries with gravitational wave data, [207] demonstrated that SBI can recover population-level parameters bypassing individual sources' fits. In exoplanetary atmospheric retrieval, [208] obtained orders-of-magnitude efficiency increase with respect to nested sampling methods, enabling large-scale model comparison studies. More broadly, SBI is flexible enough to operate both on summary statistics, such as the traditional power spectrum and bispectrum (see e.g. [209]) or more advanced ones (see e.g. [210]), and on full field-level data using realistic forward models (see e.g. [211, 212]). Complementarily, recent field-level SBI methods have also targeted the reconstruction and fast posterior sampling of cosmological initial conditions [213, 214].

Despite its appeal, SBI requires a large number of simulations to in order to train

the inference model. Recent weak-lensing studies have benchmarked full-field approaches, emphasizing both the potential optimality of field-level inference and the large simulation budgets required to achieve sufficient generalization performance [215]. This issue has also been studied in [216], which showed that the number of simulations needed for neural information extraction can be substantially larger than in currently available simulation suites. This is where the emulators introduced in Section 1.4.2 enter the picture, as they have become an essential component of modern inference pipelines, owing to their ability to generate realistic simulations at a fraction of the computational cost.

The authors of [217] highlight a critical challenge for SBI, namely its sensitivity to distributional shifts between different simulation codes using the same parameters and initial conditions (even just in the pure N-body regime). In particular, discrepancies arising from numerical schemes of cosmological N-body simulations and effective resolution can lead to out-of-distribution (OOD) behaviors and hence biased parameter inference. These effects appear to be driven by small-scale differences, to which neural networks are especially sensitive, and can sometimes be mitigated by applying appropriate smoothing to restore statistical consistency between datasets, although this comes at the expense of constraining power.

Finally, one last key aspect of modern LSS analyses relevant to this work is the joint exploitation of multiple tracers of the underlying matter distribution. Different observables probe the same large-scale structures while being affected by distinct astrophysical processes and observational systematics. As a result, their combination through cross-correlations provides a promising path to enhance the overall constraining power by reducing such systematics [218, 219]. Previous work includes the cross-correlation of different LSS tracers [220, 221, 222, 223], LSS with the Cosmic Microwave Background [224, 225, 226, 227, 228, 229, 230, 231, 232, 233], the IM of different lines [234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246], neutrinos [247], Gravitational Waves [248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266], or machine learning based multi-field analysis directly from simulations [267].

The goal of this work is to develop a proof-of-concept exploration of an SBI approach to multi-field LSS analysis, by developing a blueprint for a scalable SBI pipeline capable of inferring marginal posterior distributions for cosmological parameters, namely $\Omega_{\rm m}$ and σ_8. We focus on two complementary observational probes: galaxy number density fields and the mean brightness temperature of neutral hydrogen. By analyzing these tracers either individually or in combination, we aim to assess the gain in information when moving from single-tracer to joint analyses. We also investigate the information gain that can be expected from overcoming standard compression schemes like the power spectrum to machine-learning derived representations of full-field 3D inference. Emulators play a key role in this work, since they allow us to efficiently generate large numbers of galaxy and

HI maps starting from dark matter-only simulations, thus bypassing the need for large numbers of full-hydrodynamical runs. Our approach thus combines (i) fast approximate dark matter simulations, (ii) emulators to generate galaxy and HI fields, and (iii) SBI techniques to perform inference either on summary statistics or on latent representation of 3D fields. This allows us to systematically investigate the trade-offs between computational cost, data compression, and information retention, and to quantify gains and challenges of moving towards full-field 3D inference and multi-tracers analyses.

4.2 Methodology and results

This section outlines the methodology of this pipeline, aiming to develop a flexible SBI pipeline capable of extracting cosmological information directly from large-scale structure fields and assess the advantages and limitations of field-level SBI on multiple tracers.

We adopt a controlled and idealized setup that allows us to isolate the cosmological information content of the simulated fields. In particular, we account for cosmic variance by considering multiple realizations and using an averaging procedure over initial conditions, thus ensuring that inference is not driven by specific initial conditions. At the same time, we deliberately do not include additional sources of observational uncertainty—such as instrumental noise, survey systematics, or foreground contamination—in order to establish a clear and robust baseline of an idealized scenario for the performance of the pipeline. Our training, validation and testing sets are all self-consistently based on emulated maps, given that thoroughly testing the validity of transfer learning to the original suite (or to different ones) would require a full OOD analysis that go beyond the scope of this idealized setup work. In this sense, our analysis should be regarded as a proof-of-concept, designed to demonstrate the feasibility, potential and limitations of the proposed SBI framework, especially in the field-level and multi-tracers approaches. Building on these results, a natural next step will be to incorporate increasingly realistic observational effects and to tailor the analysis to the characteristics of ongoing and future surveys, further enhancing the applicability of our methodology.

4.2.1 Overview of pipeline

Before detailing each component of our method, here we provide a high-level description:

1. In order to obtain observable predictions for LSS, we start from approximate dark matter (DM)-only 3D maps, evolved from high redshift down to $z = 0$ with `FastPM` [89]. Maps feature a box size of either 25 or $50\,h^{-1}$Mpc; additional details are given in section 4.2.2.

2. We train a modified version of the the U-Net [268] emulator presented in [144], `HALOgen`, to generate, conditioned on the DM maps, 3D maps of two different observables: galaxy number counts density and mean brightness temperature T_b of neutral hydrogen. We refer to such emulator as `HGL` (`HALOgen`-like). For emulator training, this work takes advantage of the full-hydrodynamical simulations from the Latin Hypercube (LH) suite of the `CAMELS` Simulation Project [182] with prescriptions from the `IllustrisTNG` [183, 184] suite. Such simulations are available for a different combination of cosmological parameters ($\Omega_{\rm m}$, σ_8) and astrophysical ones ($A_{\rm SN1}, A_{\rm AGN1}, A_{\rm SN2}, A_{\rm AGN2}$), describing the relative strength and efficiency of feedback processes from Supernovae and Active Galactic Nuclei. More details on observables construction are provided in section 4.2.3, while for details on the emulator training and architecture we refer the reader to section 4.4.

3. We perform SBI on emulated galaxy/HI maps to infer (marginal) posterior distributions on cosmological parameters. We investigate two approaches: first, we perform SBI on a summary statistic of the full field, the power spectrum $P(k)$ (see top panel of figure 4.1); second, we perform SBI on a machine learning representation of the full field (see bottom panel of figure 4.1). Within each, we compare the constraints on cosmology that can be achieved from one single tracer (galaxy or HI) or from simultaneously considering both (different input lines in figure 4.1, bottom). See section 4.2.4 for full details.

Figure 4.1 provides a schematic overview of our pipeline, with each step described in detail below. In this work, all observables are considered at redshift $z = 0$, working in the plane-parallel approximation and assuming the box redshift to be constant. Extending the analysis to higher z is planned for future works.

4.2.2 Dark matter approximate maps

We train the `HGL` emulator to produce 3D galaxy/HI fields starting from an approximate, $z = 0$ 3D DM map. For the purpose of constructing a suitable training set for the emulator, we make use of `FastPM` to generate approximate DM maps at small computational cost, and we pair them with galaxy/HI 3D maps obtained from full hydro simulations from `CAMELS` for the same cosmological parameters and initial conditions (further details on galaxy emulator train/validation sets in section 4.4). `FastPM` takes as inputs the cosmological parameters $\{\Omega_{\rm m}, \sigma_8\}$ (the latter in the form of the normalization of the input power spectrum), the linear total matter power spectrum, and a seed for the generation of the linear density map at high redshift.

The `FastPM` random number generator is compatible with the one used for `CAMELS` (part of `N-GenIC` [160]), therefore, for the training set, we simply feed `FastPM` the seeds

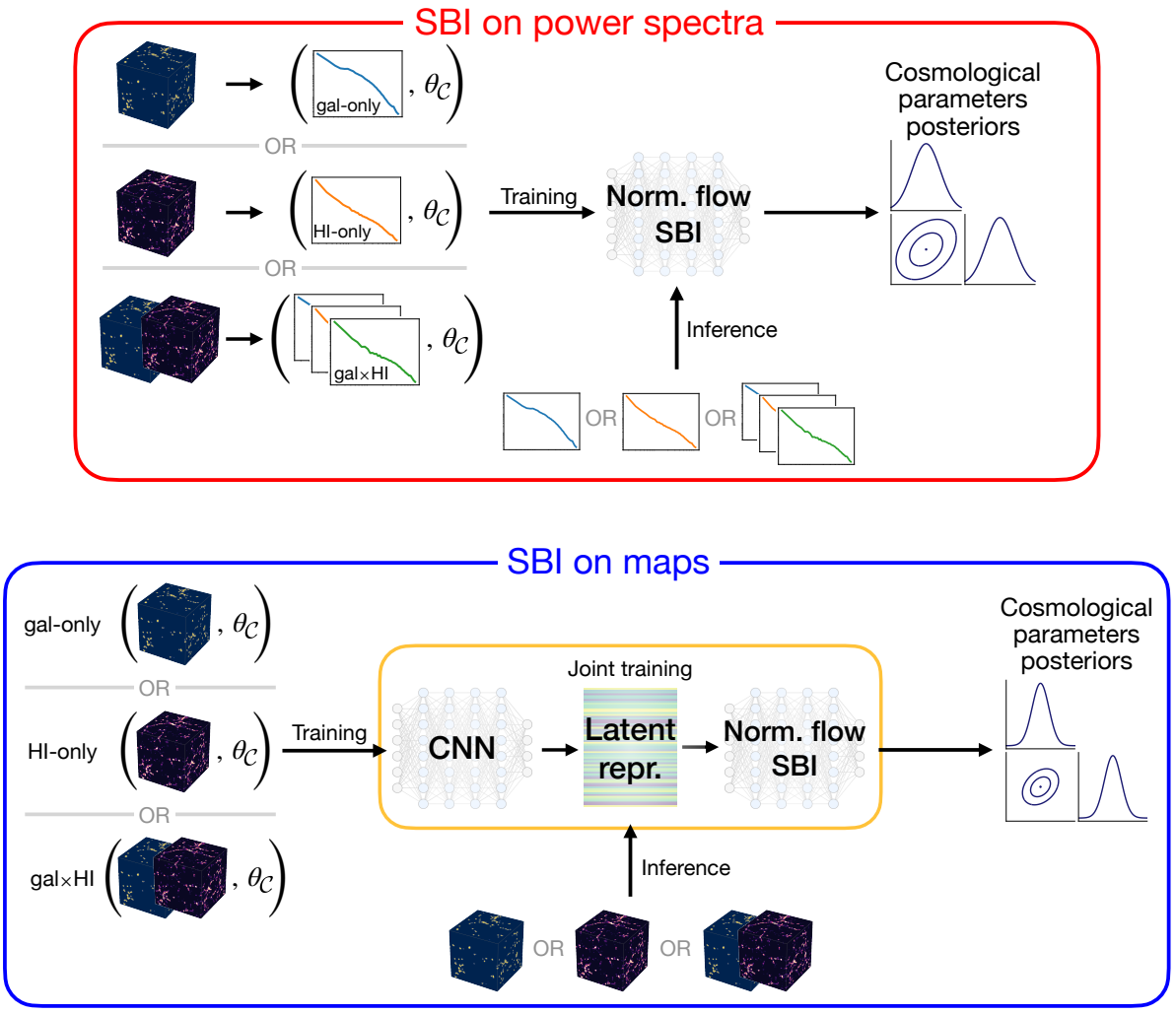


Figure 4.1: *Top*: sketch of the SBI on power spectra pipeline. Power spectra are extracted from the emulated galaxy and/or HI maps, so that pairs of power spectra and cosmological parameters $\theta_{\mathcal{C}} = \{\Omega_{\rm m}, \sigma_8\}$ constitute the SBI training set (observables fields are reduced to power spectra). Once trained, the SBI flow is able to produce posteriors on $\theta_{\mathcal{C}}$ for validation power spectra. *Bottom*: sketch for SBI on maps pipeline. Pairs of emulated full maps and $\theta_{\mathcal{C}}$ are passed for joint training to a Convolutional Neural Network (CNN) which extracts a latent representation of the data, which is then used to train the SBI normalizing flow.

provided in the `CAMELS` dataset, as well as the linear power spectrum also provided in the `CAMELS` simulation suite data. On the contrary, when emulating observables from new DM maps (i.e. when training and testing the SBI model), we generate the linear power spectrum using the Boltzmann solver `CLASS` [41] and we feed `FastPM` arbitrary seeds.

We generate DM cubes with both $L = 25\,h^{-1}$Mpc and $L = 50\,h^{-1}$Mpc sides (we discuss this choice more in detail below), with the following properties:

- $L = 25\,h^{-1}$Mpc: `FastPM` is run with 128^3 particles on a 128^3 mesh and maps are saved with a mesh size of 64^3;

- $L = 50\,h^{-1}$Mpc: `FastPM` is run with 256^3 particles on a 256^3 mesh and maps are saved with a mesh size of 128^3.

With these specifications, maps of different sizes have the same resolution. In each case, we generate a linear density map with twice the resolution of the mesh used for the simulator in order to improve the accuracy on small scales. We then generate initial conditions at redshift $z = 9$ using `2LPT` [269] and run 10 `FastPM` steps down to $z = 0$ using a force resolution $B = 2$, with an overall compute time of ~ 0.3 CPU hours per simulation (~ 35 seconds on one node of the Leonardo cluster of CINECA).

4.2.3 Galaxies and HI maps

The first tracer we consider consists of galaxy number counts density maps. To generate these maps, we carry out the following steps: we choose the full-hydrodynamical snapshots from the `CAMELS` LH suite with comoving volume $L^3 = (25\,h^{-1}\text{Mpc})^3$ and $N_{\text{part}} = 256^3$ DM particles and gas elements, at $z = 0$. We apply a mask selecting sub-halos with mass larger than a threshold[1] of $10^9 M_\odot/h$ and non-zero stellar mass, and generate the density field with the cloud-in-cell (CIC) mass-assignment scheme on a grid of size $N_{\text{grid}} = 64$, further applying a smoothing with a Gaussian kernel of radius $R = 0.2\,h^{-1}$Mpc (following [144]) and, to help training, clipping low values of the density field to a minimum of 10^{-4}.

Maps of the second tracer, HI temperature, are obtained from the correspoding full hydrodynamic snapshots with the same L^3 and N_{part} values, at $z = 0$. We extract the neutral hydrogen fraction $x_{\text{HI},i}$ and gas mass $m_{g,i}$ $[M_\odot]$ for each gas particle i in the simulation. The corresponding neutral hydrogen density is calculated as

$$\rho_{\text{HI},i} = m_{g,i} X x_{\text{HI},i} \left(\frac{N_{\text{part}}}{a \times L}\right)^3 \left[\frac{\text{kg}}{\text{m}^3}\right] , \tag{4.1}$$

where $X = 0.76$ is the hydrogen mass fraction as set in `CAMELS` and a is the scale factor. We set $x_{\text{HI},i} = 1$ for particles with non-zero star formation rate, following the prescription adopted in `CAMELS` [271]. This approximation assumes complete self-shielding and reproduces key observational constraints, including the HI abundance and the column density distribution of damped Lyman-α systems [270]. We then interpolate $\rho_{\text{HI},i}$ onto a continuous field ρ_{HI} using the CIC scheme and $N_{\text{grid}} = 64$. The 21 cm brightness temperature in the post-reionization regime is approximated as [270]

$$T_b(\mathbf{x}) = 189h \left(\frac{H_0(1+z)^2}{H(z)}\right) \frac{\rho_{\text{HI}}(\mathbf{x})}{\rho_c} \,[\text{mK}] , \tag{4.2}$$

[1] This threshold is chosen for two reasons: first, we would like to have numerically resolved haloes and this lower limit guarantees at least 100 dark matter particles per halo; second, we would like to use all the haloes which are likely to host HI according to more sophisticated hydro sims like [270].

where $\rho_{\rm HI}(\mathbf{x})$ is the interpolated neutral hydrogen density at position $\mathbf{x}$, $H(z)$ is the Hubble–Lemaître parameter, $h = 0.6711$ is the dimensionless Hubble–Lemaître constant (fixed at `CAMELS` fiducial value) and $\rho_c = 9.47 \cdot 10^{-27} kg/m^3$ is the critical density. This field is also smoothed with a $R = 0.2\,h^{-1}$Mpc Gaussian kernel.

The process of predicting such 3D galaxy and HI fields from their approximate DM counterpart can be thought of an image to image translation task. To achieve this while preserving spatial correlations and structure we make use of `HGL`. While the original `HALOgen` [144] was designed to work for fixed cosmological parameters, here we introduce structural modifications to generalize emulated fields for different cosmological $\theta_{\mathcal{C}} = \{\Omega_{\rm m}, \sigma_8\}$ and astrophysical $\theta_{\mathcal{A}} = \{\mathrm{A_{SN1}}, \mathrm{A_{AGN1}}, \mathrm{A_{SN2}}, \mathrm{A_{AGN2}}\}$ parameters (defined as in `CAMELS`).

One limitation of using relatively small boxes is the power spectrum sensitivity to cosmic variance and finite-volume effects (mentioned repeatedly in Chapter 3). Indeed, we found that, for $25\,h^{-1}$Mpc boxes, realizations at fixed physical parameters and generated with different random seeds can yield noticeably different power spectra (creating a degeneracy between cosmology and the particular realization of the density field); for this reason, we adopt boxes of size $50\,h^{-1}$Mpc to mitigate this effect and improve the sampling of large-scale modes. However, training `HGL` for $50\,h^{-1}$Mpc boxes is not feasible for the purpose of this work within the `CAMELS` simulations, since the LH set is not available for a box size of $50\,h^{-1}$Mpc, and the available SB35 set for such box size varies 30 astrophysical parameters at the same time (instead of 4), making the emulation task significantly more difficult. For these reasons, we train `HGL` using the LH simulations at $25\,h^{-1}$Mpc, and then use it to emulate maps at $50\,h^{-1}$Mpc. Note that emulating maps of different size is correct only wheen keeping the same resolution, motivating the choice of number of particles and grid size for the `FastPM` simulations mentioned above. Details about the emulator architecture and training are provided in section 4.4.

4.2.4 Simulation-Based Inference with Neural Posterior Estimation

We use the galaxy and HI emulated maps to perform SBI in order to obtain posterior distributions of the cosmological parameters $\theta_{\mathcal{C}}$; this is done by either fixing the 4 astrophysical parameters $\theta_{\mathcal{A}}$ to fiducial values $\{\mathrm{A_{SN1}} = \mathrm{A_{AGN1}} = \mathrm{A_{SN2}} = \mathrm{A_{AGN2}} = 1\}$ or by marginalizing over them. SBI is trained on newly generated maps, in terms of both seeds and parameters values, obtained via the emulator from uniform priors as follows: $\Omega_{\rm m} \in [0.2, 0.4]$ and $\sigma_8 \in [0.7, 0.9]$ and, for the free astrophysics scenario, $\mathrm{A_{SN1}} \in [0.7, 1.5], \mathrm{A_{AGN1}} \in [0.7, 1.5], \mathrm{A_{SN2}} \in [0.5, 2.0], \mathrm{A_{AGN2}} \in [0.5, 2.0]$.

We perform SBI with Neural Posterior Estimation (NPE) [272, 273, 274, 275], in which

a neural density estimator is trained to approximate the conditional posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{d})$ directly from simulated parameters-data pairs, $(\boldsymbol{\theta}, \mathbf{d})$. In this work, the posterior is modeled using a normalizing flow [276], i.e. an expressive, invertible transformation that maps a simple base distribution (typically a multivariate Gaussian) into a complex target distribution through a sequence of learned bijective transformations. NPE training maximizes the log-probability of sampled parameters given simulated data, i.e. minimizing the negative log-likelihood $\mathcal{L}(\phi) = -\mathbb{E}_{p(\boldsymbol{\theta},\mathbf{d})}[\log q_\phi(\boldsymbol{\theta} \mid \mathbf{d})]$, where $q_\phi(\boldsymbol{\theta} \mid \mathbf{d})$ denotes a parametric approximation of the true posterior and ϕ are the normalizing flow parameters (see e.g. [272, 274] for further details).

Specifically, we adopt a Masked Autoregressive Flow (MAF) [277], a type of normalizing flow that uses autoregressive neural networks to transform a simple probability distribution into a complex one by masking weights to ensure each dimension only depends on previous ones. The main advantage of MAFs is that they provide highly efficient density estimation, allowing for relatively quick training.

As anticipated in section 4.2.1, we perform SBI on two different types of data:

- **SBI on power spectrum** (case $\mathcal{P}$): we perform SBI on the power spectra $P(k)$ extracted from the emulated galaxy/HI maps, as a baseline case of the achievable constraints within an SBI framework with this summary statistic alone. Firstly, we perform SBI looking only at one tracer (either galaxy or HI); then we take into account the full cross-correlation g×HI information, given by both the auto-correlation power spectra $P_k^{\rm g}$ and $P_k^{\rm HI}$ and the cross-correlation power spectrum $P_k^{\rm g\times HI}$. For the sake of simplifying the notation, in what follows these three cases are referred to as, respectively, $\mathcal{P}^{\rm g}$, $\mathcal{P}^{\rm HI}$ and $\mathcal{P}^{\rm g\times HI}$. Full details on the methodology are provided later in this section.

- **SBI on latent representations of full maps** (case $\mathcal{M}$): the entire density field is used here, to inform a latent representation of the full galaxy/HI maps, extracted by a Convolutional Neural Network (CNN) attached to the SBI flow for joint training. As in the previous scenario, we first work on one single field and then feed both of them to the network jointly. Such cases are referred to as $\mathcal{M}^{\rm g}$, $\mathcal{M}^{\rm HI}$ and $\mathcal{M}^{\rm g\times HI}$. We explore both the case where maps are flattened (2D$\mathcal{M}$) by averaging along a third axis, and a full three-dimensional scenario (3D$\mathcal{M}$), which is more challenging from a memory and training perspective, while providing in principle more information. Full details are provided later in this section.

Table 4.1 details the different SBI setups explored and summarizes the nomenclature that identifies each configuration.

Table 4.1: Summary of the different SBI setups explored in this work grouped by category, and corresponding abbreviations. For example: SBI on power spectrum performed with fixed astrophysics and galaxies as single tracer is referred to as $\mathcal{P}^{\rm g}_{\rm fixedA}$ and SBI on maps performed on 3D cubes of both observables, with free astrophysics is referred to as $3D\mathcal{M}^{\rm g\times HI}$.

CATEGORY	**OPTIONS**
Data representation	$P(k)$: $\mathcal{P}$ representation of full-field maps: 2D$\mathcal{M}$ or 3D$\mathcal{M}$
Astrophysics	fixed: subscript "fixedA" free: no subscript
tracers	gal-only: apex "g" HI-only: apex "HI" galaxies×HI: apex "g×HI"

The inference pipeline is implemented using the `sbi` library [278]. We test different hyper-parameters configurations; the chosen values are specified in each corresponding subsection.

Case $\mathcal{P}$: SBI on Power Spectrum

For the SBI analysis performed on power spectrum (case $\mathcal{P}$), we compress the three-dimensional information of the emulated galaxy/HI field into its isotropic power spectrum $P(k)^{\rm X}$, with $X \in \{\rm g, HI\}$. We compute it by means of the `Pylians` library [162]: starting from the density field defined on a regular grid, we perform a Fast Fourier Transform to obtain the Fourier modes and estimate the spherically-averaged power spectrum, binning modes in shells of constant wavenumber k: $P(k)^{\rm X} = \langle|\delta^{\rm X}(k)|^2\rangle$, where $\delta^{\rm X}(k)$ is the Fourier transform of the galaxy number density/HI mean brightness temperature field (both normalized by their mean). Similarly, for the cross-correlation case we compute $P(k)^{{\rm X},Y} = \langle\delta^{\rm X}(k)\delta^{\rm Y^*}(k)\rangle$ with $\rm X \neq Y$. Computations are performed over the range of scales allowed by the box size and grid resolution.

We consider power spectra extracted from boxes of size 50 h^{-1}Mpc, in the scales range $k \in [0.18, 4.7]\, h$/Mpc. For each combination of physical parameters we average the power spectra obtained from 10 independent realizations with different initial seeds. This procedure reduces the impact of realization-to-realization fluctuations and is effectively equivalent to probing a larger survey volume, as it increases the number of independent modes contributing to the estimated power spectrum. We remind the reader that in our simplified setup no measurement noise is added to the simulated data. In this regard, the power-spectrum-related uncertainties treatment includes only the aforementioned contribution from cosmic variance.

To build the SBI training set for the $\mathcal{P}$ case, we emulate 8×10^4 galaxy/HI maps drawing physical parameters from their prior ranges and producing 10 emulations (each with different initial conditions) for each parameter set value, over which we average the power spectra. This produces 8×10^3 averaged power spectra. The validation set

is obtained as 20% of the training set, randomly sampled (after averaging over initial conditions). The SBI test set is made of 200 averaged (each over 10 initial conditions) power spectra, all drawn from the fiducial ΛCDM scenario with $\theta_{\mathcal{C}} = \{\Omega_{\rm m} = 0.315,\ \sigma_8 = 0.811\}$ from Planck fiducial values [5] and either with $\theta_{\mathcal{A}} = \{{\rm A_{SN1}} = {\rm A_{AGN1}} = {\rm A_{SN2}} = {\rm A_{AGN2}} = 1\}$ (for the fixed astrophysics baseline case) or marginalizing over different astrophysics values within their uniform priors. After the averaging and before passing the power spectra to the SBI flow, we move to $\log_{10}$ space and normalize the sets with standard scaling.

For the cross-correlation cases, $\mathcal{P}^{\rm g\times HI}_{\rm fixedA}$ and $\mathcal{P}^{\rm g\times HI}$, we pass to the SBI flow a concatenation of both the auto-spectra and the cross-spectrum, after they have been re-normalized all together: $\{P(k)^{\rm g}, P(k)^{\rm HI}, P(k)^{\rm g,HI}\}$.

The MAF network adopted to model the posterior distribution is configured with 5 transforms and 128 hidden features per layer. Training is performed with a batch size of 128 and a learning rate of 10^{-3}. We train with early stopping applied if the validation loss does not improve for 120 consecutive epochs. Such hyper-parameters configuration are chosen after exploring several different combinations.

Section 4.3 provides results for posteriors on $\{\Omega_{\rm m}, \sigma_8\}$ and calibration plots both fixing the astrophysics to fiducial values and letting it free, marginalizing over the four astrophysical parameters.

Case $\mathcal{M}$: SBI on latent representations from maps

Field-level SBI relies on a CNN to compress the input emulated maps into a one dimensional latent space, which is then passed to the normalizing flow. The CNN encoder and the flow are trained jointly in order to obtain a physically informative latent space. Compressing the maps by means of a separate autoencoder guided by a reconstruction loss would not necessarily preserve the most informative features for the physical parameters of interest.

In a computationally less demanding configuration, we flatten the full maps to 2 dimensions by averaging over the third axis, and extract the latent space from them. We refer to this scenario as 2D$\mathcal{M}$ case. We also explore the increase of information provided by the entire three-dimensional map. This scenario, referred to as 3D$\mathcal{M}$ case, is ideally expected to provide the highest amount of information, but at a significantly higher computational cost.

The SBI train set is here composed of 8×10^4 (3×10^4) maps of box size $50\,h^{-1}$Mpc and $N_{\rm grid} = 128$ for the 2D (3D) case, the 3D set being smaller for computational costs reasons. To account for cosmic variance and finite box effects, 10 different initial conditions are adopted for each parameter value, as was done in the $\mathcal{P}$ case. The test set is instead made of 200 ΛCDM-Planck realizations, with different never-seen-before seeds, same cosmology

$\theta_{\mathcal{C}} = \{\Omega_{\rm m} = 0.315,\ \sigma_8 = 0.811\}$ [5] and either with fixed astrophysics ($\theta_{\mathcal{A}} = \{\mathrm{A_{SN1}} = \mathrm{A_{AGN1}} = \mathrm{A_{SN2}} = \mathrm{A_{AGN2}} = 1\}$) or marginalizing over astrophysical effects within their priors.

The CNN encoder is implemented in either 2D or 3D depending on the dimensionality of the input maps, with the same architecture applied in both cases. It is composed of two layers with 32 and 64 filters, respectively, using kernel size 3 and stride 2 in each layer and a ReLU activation function. We adopt group normalization, with 4 and 8 groups in the first and second layers, respectively. In the single tracer cases, $\mathcal{M}^{\rm g}$ and $\mathcal{M}^{\rm HI}$, the network is characterized by only one channel, whereas in the cross-correlation case, $\mathcal{M}^{\rm g\times HI}$, we have one channel for each tracer, for a total of two. The final latent space is obtained via average pooling over the spatial dimensions and has a single dimension of length 64 for all cases.

Similarly to the $\mathcal{P}$ case, we model the posterior using a MAF with 5 transforms and 128 hidden features per layer. The network is trained with a learning rate of 10^{-3}, a batch size of 128 (16) with early stopping if the validation loss does not improve for 120 (35) epochs for the 2D (3D) case. We reserve 20% of the simulations for validation. Due to the extreme computational costs of full-field SBI, especially for the 3D case, we could not perform parameters tuning as extensively as in the $\mathcal{P}$ case. We chose, instead, configurations analogue to the best $\mathcal{P}$ case ones. Table 4.2 summarizes the architecture for the $\mathcal{M}$ cases.

Table 4.2: Architecture of the SBI model for 3D field-level inference (model 3D$\mathcal{M}$). The input maps are compressed into a 64-dimensional latent representation via a convolutional encoder, and the posterior over parameters is inferred using NPE with a MAF. Note that $d = 2$ (3) for 2D (3D) field-level inference.

Stage	**Representation**	**Channels**	**Operation**
Input	128^d maps	1 or 2	Input fields
Encoder	$(128/2)^d$	32	Conv (k=3, s=2) + GN (g=4) + ReLU
	$(128/4)^d$	64	Conv (k=3, s=2) + GN (g=8) + ReLU
Latent	$\mathbb{R}^{64}$	–	Global mean pooling
Flow	$\mathbb{R}^{64}$	–	MAF (5 transforms, 128 hidden features)
Output	$p(\boldsymbol{\theta} \mid \mathbf{d})$	–	Approximate posterior distribution

To generate the results below, the $\mathcal{P}$ cases were trained for around 2 minutes on a Apple McBookPro M4 CPU, the 2D$\mathcal{M}$ took about 40 minutes on a single A100 GPU whereas training the 3D$\mathcal{M}$ cases required one full week on the same single A100 GPU (although preliminary posteriors near convergence were already available after a runtime of ~ 2 days). In all cases, inference requires only a few seconds on the Apple McBookPro M4 CPU. In terms of disc usage for the $50\,h^{-1}$Mpc case, the full train set for the $\mathcal{P}^{\rm g\times HI}$ scenario occupies a mere $\sim 1\,$MB ($8 \times 10^3 \times 3$ average power spectra), the one for the

2D$\mathcal{M}^{\mathrm{g\times HI}}$ case $\sim 10\,\mathrm{GB}$ ($8\times10^4\times2$ maps) and finally $\sim 470\,\mathrm{GB}$ for 3D$\mathcal{M}^{\mathrm{g\times HI}}$ ($3\times10^4\times2$ cubes).

4.3 Results

4.3.1 Statistical precision

Figure 4.3[2] shows posteriors on $\{\Omega_\mathrm{m}, \sigma_8\}$ for the 200 realizations in the test set (at fixed ΛCDM Planck ground truth parameters) for the baseline fixed-astrophysics scenario, giving an overview of the gains reachable with field-level SBI. The following cases are shown: $\mathcal{P}^{\mathrm{g\times HI}}_{\mathrm{fixedA}}$, 2D$\mathcal{M}^{\mathrm{g\times HI}}_{\mathrm{fixedA}}$ and 3D$\mathcal{M}^{\mathrm{g\times HI}}_{\mathrm{fixedA}}$. The light lines indicate contours for each test set realization, whereas solid lines provide the mean posterior, averaged overall all realizations. Although the 2D$\mathcal{M}^{\mathrm{g\times HI}}_{\mathrm{fixedA}}$ case does not significantly improve constraining power, the 3D$\mathcal{M}^{\mathrm{g\times HI}}_{\mathrm{fixedA}}$ case provides a significant gain, especially for σ_8. An even clearer improvement can be seen when marginalizing over the astrophysical parameters: figure 4.4 displays significant improvements on both $\{\Omega_\mathrm{m}, \sigma_8\}$ when moving to 3D field-level (case 3D$\mathcal{M}^{\mathrm{g\times HI}}$). In figure 4.5 we display posteriors for the free astrophysics 3D$\mathcal{M}^{\mathrm{g\times HI}}$ case, comparing it with the baseline 3D$\mathcal{M}^{\mathrm{g\times HI}}_{\mathrm{fixedA}}$ scenario. Although marginalization over astrophysical parameters reduces the precision of marginal posteriors by about a factor ~ 2 for both Ω_m and σ_8, additional cosmological information is still clearly retained in the field-level approach. This is different for the $\mathcal{P}^{\mathrm{g\times HI}}$ case previously shown in figure 4.4, where it can be noticed that limiting the analysis to the power spectrum provides very wide posteriors, especially on σ_8 where the posterior spans nearly the whole prior range. Including information beyond the power spectrum seems therefore crucial to retain cosmological information while marginalizing over astrophysics. Figure 4.6 (fixed astrophysics) and figure 4.7 (free astrophysics) also help highlighting the improvement obtainable in terms of posteriors predictivity when moving from single tracer to multi-tracers analysis. In particular, we notice how including HI information produces a dramatic improvement in constraints compared to galaxy tracers only.

For each combination of tracers and choice of data summary, we quantify the resulting statistical precision on the two cosmological parameters of interest by means of a Figure of Merit (FoM) defined as $\mathrm{FoM} = 1/\mathrm{A}_{68}$, where A_{68} is the area within the 68% joint credible region for σ_8 and Ω_m (marginalized over astrophysical parameters where appropriate). Table 4.3 shows the median Figure of Merit (FoM_m) computed from the the 200 posteriors in the test set. When moving from $\mathcal{P}^{\mathrm{g\times HI}}$ to 3D$\mathcal{M}^{\mathrm{g\times HI}}$, FoM_m increases by a factor ~ 3 for both the fixed and free astrophysics configurations, giving a first indication of the gains achievable by a 3D field-level analysis. Comparison with the 2D$\mathcal{M}^{\mathrm{g\times HI}}$ cases shows that

[2] For readability, all figures of this section are collected at its end.

flatteninig the cubes to 2D maps does not yield comparable increase. Furthermore, both in the $\mathcal{P}$ and $\mathcal{M}$ scenarios, the multi-tracers cases are those yielding the highest FoM$_m$, which increases with respect to the single-tracer cases by factors between ~ 2 and up to 7. This demonstrates a non-negligible improvement in the precision achievable by the multi-tracers approach. Accuracy and calibration of the posteriors are addressed below.

Data summary	**Tracer**	**FoM$_m$ (Fixed A.)**	**FoM$_m$ (Free A.)**
$\mathcal{P}$	g	0.65	0.33
$\mathcal{P}$	HI	2.5	0.27
$\mathcal{P}$	g×HI	4.7	1.0
2D$\mathcal{M}$	g×HI	5.2	1.2
3D$\mathcal{M}$	g	2.4	0.47
3D$\mathcal{M}$	HI	5.5	2.2
3D$\mathcal{M}$	g×HI	12	3.4

Table 4.3: Median Figure of Merit FoM$_m$ across test sets, for the explored SBI scenarios. Values are normalized by 3.23×10^2, corresponding to the FoM$_m$ of $\mathcal{P}^{\rm g\times HI}$ case.

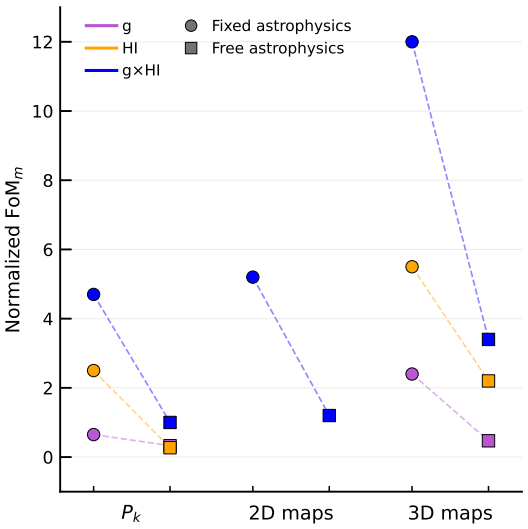


Figure 4.2: Visualization of table 4.3 data.

To further investigate the constraining power of the various scenarios, we show in figures 4.8, 4.9, 4.10 and 4.11 the distribution of marginal posterior means over the 200 test realizations (top panel) and their median widths (bottom panel). Figures 4.8 and 4.9 display the results for fixed astrophysics for $\Omega_{\rm m}$ and σ_8, respectively: the 3D$\mathcal{M}^{\rm g\times HI}_{\rm fixedA}$ case delivers unbiased and narrower posteriors with respect to the $\mathcal{P}^{\rm g\times HI}_{\rm fixedA}$ case. This is especially prominent for σ_8, whose constraints improve by a factor ~ 2 (while $\Omega_{\rm m}$ is improved only mildly). It is worth noticing that the 2D$\mathcal{M}^{\rm g\times HI}_{\rm fixedA}$ case, instead, does not yield a significant improvement, especially for $\Omega_{\rm m}$, where in fact we observe a degradation of constraints with respect to the $\mathcal{P}^{\rm g\times HI}_{\rm fixedA}$ case. This indicates that, at least under the current setup for the latent representation, the three-dimensionality of the field-level maps is required to retain additional cosmological information.

Focusing the attention on the improvements from combining both tracers, figures 4.8 and 4.9 also show that posteriors for the 3D$\mathcal{M}^{\rm g\times HI}_{\rm fixedA}$ case are narrower by a factor ~ 4 for $\Omega_{\rm m}$ and $\sim 4-5$ for σ_8 with respect to the 3D$\mathcal{M}^{\rm g}_{\rm fixedA}$ case and narrower by about $\sim 40\%$ for $\Omega_{\rm m}$ with respect to the 3D$\mathcal{M}^{\rm HI}_{\rm fixedA}$ case, indicating a clear advantage from multi-tracer combination. This is most likely due to the fact that galaxy maps are significantly sparser, thereby affected by shot-noise, and built from halos spanning a wide range of masses, each with different bias factors, which effectively mixes objects tracing the underlying DM dis-

tribution in different ways. In contrast, HI intensity maps trace the matter field as a smoother continuous observable with a more straightforward dependence on the underlying matter density. Furthermore the relatively small box size used in this study prevents us to fully exploit the capabilities of galaxy clustering at linear scales. The main message of the exercise performed in this work is that the HI field retains more information, at non-linear level, compared to the galaxy field, and our SBI-based framework is able to extract this information. Indeed, this should not be interpreted as a limitation of galaxy clustering analyses per se, nor as a contradiction with the success of traditional large-scale structure cosmology based on it, since they typically rely on physically motivated bias models (e.g. within the effective field theory framework), large survey volumes, and additional observables such as redshift-space distortions or higher-order statistics, all of which help break degeneracies and recover cosmological information. A more thorough assessment of the relative merits of each tracer would require consideration of the different noise properties of the observations and systematic effects, whose impact depends on the specific modeling assumptions and observational context, which are beyond the scope of this work. Figures 4.10 and 4.11 show the same analysis but in the case where astrophysical parameters are marginalized over. In this case, the case $3\mathrm{D}\mathcal{M}^{\mathrm{g}\times\mathrm{HI}}$ yields less biased posteriors with respect to $\mathcal{P}^{\mathrm{g}\times\mathrm{HI}}$, and clearly narrower for both cosmological parameters: a $\sim 30\%$ improvement for Ω_{m} and $\sim 50\%$ for σ_8, confirming the trends just discussed for the fixed astrophysics scenario.

Using of 200 independent test realizations allows us to assess the stability of the inference across different initial conditions. The dispersion of posterior means (figures 4.8, 4.9, 4.10, 4.11) indicates that the $3\mathrm{D}\mathcal{M}$ approach not only yields tighter constraints than the $\mathcal{P}$ scenario, but also reduces realization-to-realization scatter, suggesting that this method is less sensitive to cosmic variance, effectively extracting more robust information from each individual map.

4.3.2 Bias and calibration

In order to quantify the bias in the posterior inference of the ΛCDM-Planck test set, in table 4.4 we provide relative percentage values of the distance between the ground truth value and the median value of posterior samples from all realizations in the test set (obtained from 1000 samples for each of the 200 realizations, for fixed ΛCDM fiducial parameters values), for the scenarios explored in this work. The scenarios presenting the smallest bias are, most of the times, the $3\mathrm{D}\mathcal{M}$ cases. Nonetheless, bias trends are less extreme than the FoM_m trends reported in table 4.3 with bias values never exceeding $\sim 1.5\%$, well below statistical uncertainties.

Figure 4.13 displays the same PP-plots for the ΛCDM-Planck test set under the free astrophysics scenario, where we observe a marked improvement in calibration for the

Data summary	Tracer	Ω_m bias [%]		σ_8 bias [%]	
		Fixed A.	**Free A.**	**Fixed A.**	**Free A.**
$\mathcal{P}$	g	1.5	0.98	0.21	1.0
$\mathcal{P}$	HI	-0.23	1.0	0.64	-0.42
$\mathcal{P}$	g×HI	0.042	-0.19	1.0	0.23
2D$\mathcal{M}$	g×HI	-0.72	0.75	0.73	-0.15
3D$\mathcal{M}$	g	0.19	-1.1	-0.49	0.48
3D$\mathcal{M}$	HI	-0.38	0.23	-0.25	-0.79
3D$\mathcal{M}$	g×HI	0.084	-0.38	-0.076	0.25

Table 4.4: Bias on Ω_m and σ_8 across test sets in [%], for the explored SBI scenarios. We report the percentage difference between ground truth value and median value of the posterior samples across the test set.

2D$\mathcal{M}^{\mathrm{g\times HI}}$ scenario but a worsening for the 3D$\mathcal{M}^{\mathrm{g\times HI}}$ case with respect to the fixed astrophysics setting.

Finally, to check the calibration of our SBI pipeline across parameter space, we illustrate, in figure 4.14, the true versus marginally inferred values of cosmological parameters for the free astrophysics case 3D$\mathcal{M}^{\mathrm{g\times HI}}$, for value of $\theta_{\mathcal{A}}$ and $\theta_{\mathcal{C}}$ sampled from their priors. We also display the PP plots for this same set – this is a Bayesian PP plot, i.e., showing calibration when parameters are sampled from their respective priors. We note that the calibration in this case is compatible with the 95% binomial uncertainty essentially across the whole range, showing good calibration.

4.3.3 Overall remarks

Overall, these results indicate two main trends. First, the transition from power spectrum-based inference (model $\mathcal{P}$) to field-level representation produces the largest gains in precision when exploiting the full 3-dimensionality of the data (model 3D$\mathcal{M}$). Combining the two tracers at the power spectrum level (model $\mathcal{P}^{\mathrm{g\times HI}}$) captures only part of the available information, while the 3D$\mathcal{M}$ configurations recover a substantially larger fraction of the cosmological information, retaining well-calibrated constraining power even when marginalizing over astrophysical uncertainties. Such gain can be interpreted in terms of the additional information encoded in higher-order, non-Gaussian statistics: while the power spectrum captures only two-point statistics, the latent space representations of the full field retains non-Gaussian features generated by non-linear gravitational evolution and tracers bias. These features are particularly sensitive to σ_8, which controls the amplitude of fluctuations and therefore the strength of mode coupling. The CNN-based compression

is able to extract part of this information, whereas it is irreversibly lost when reducing the data to the power spectrum. The lack of improvement observed in the 2D$\mathcal{M}$ case also suggests that simple dimensionality reduction through projection leads to a substantial loss of information: collapsing along one axis removes line-of-sight structures which appear to carry significant cosmological information. Once again, this highlights that the gain of field-level SBI is specifically due to preserving the full three-dimensional structure, although this comes at the expense of computational cost. Second, moving from single tracer to multi-tracers analysis yields tighter and less biased constraints. This gain can be interpreted as a combination of increased effective signal-to-noise and partial breaking of degeneracies, as the two tracers probe the same underlying matter distribution with different biases and realization noise: the inclusion of cross-correlation information helps disentangle cosmological parameters from tracer-specific effects.

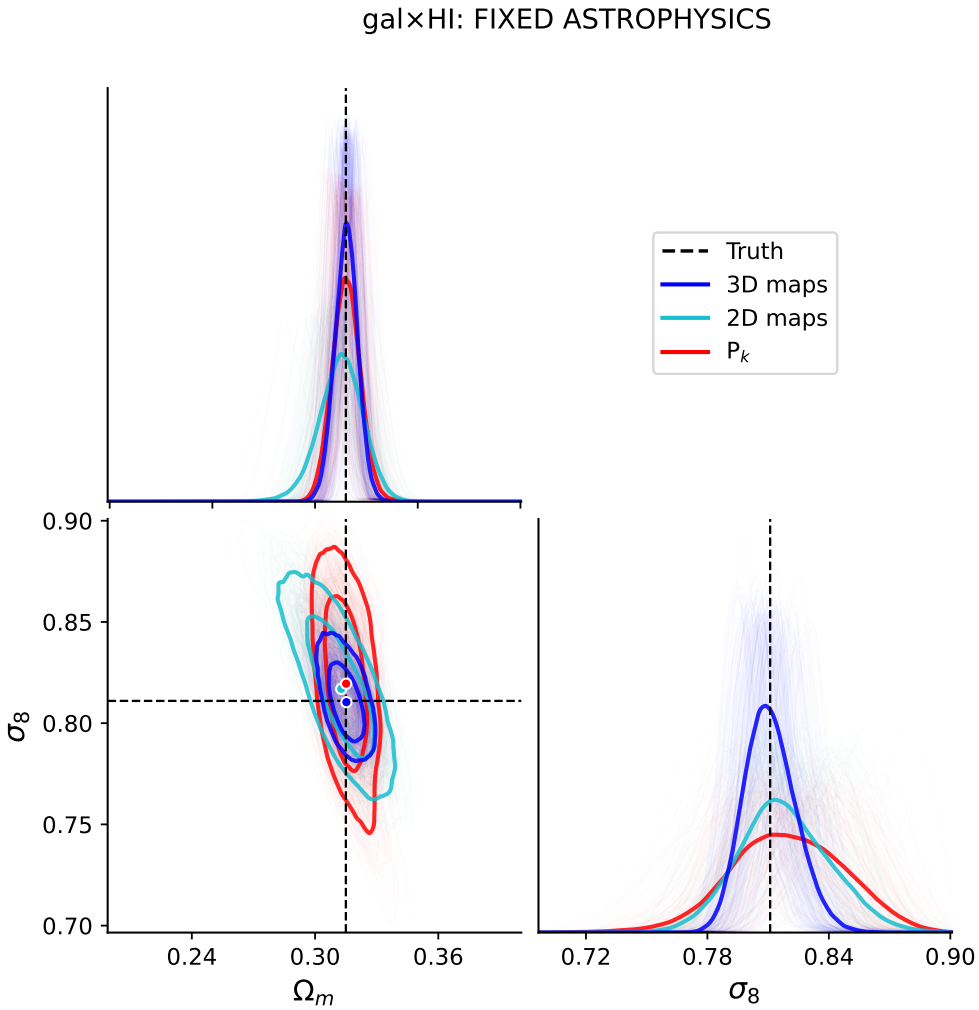


Figure 4.3: Posteriors on $\{\Omega_{\rm m}, \sigma_8\}$ for $\mathcal{P}^{\rm g\times HI}_{\rm fixedA}$ (red) vs $\mathcal{M}^{\rm g\times HI}_{\rm fixedA}$ (2D, light blue, and 3D, blue) for 200 ΛCDM-Planck realizations test set (ground truth in black). Performing inference on the full 3D maps significantly improves precision. Dots in the lower left plot represent the median values computed from all the posteriors samples.

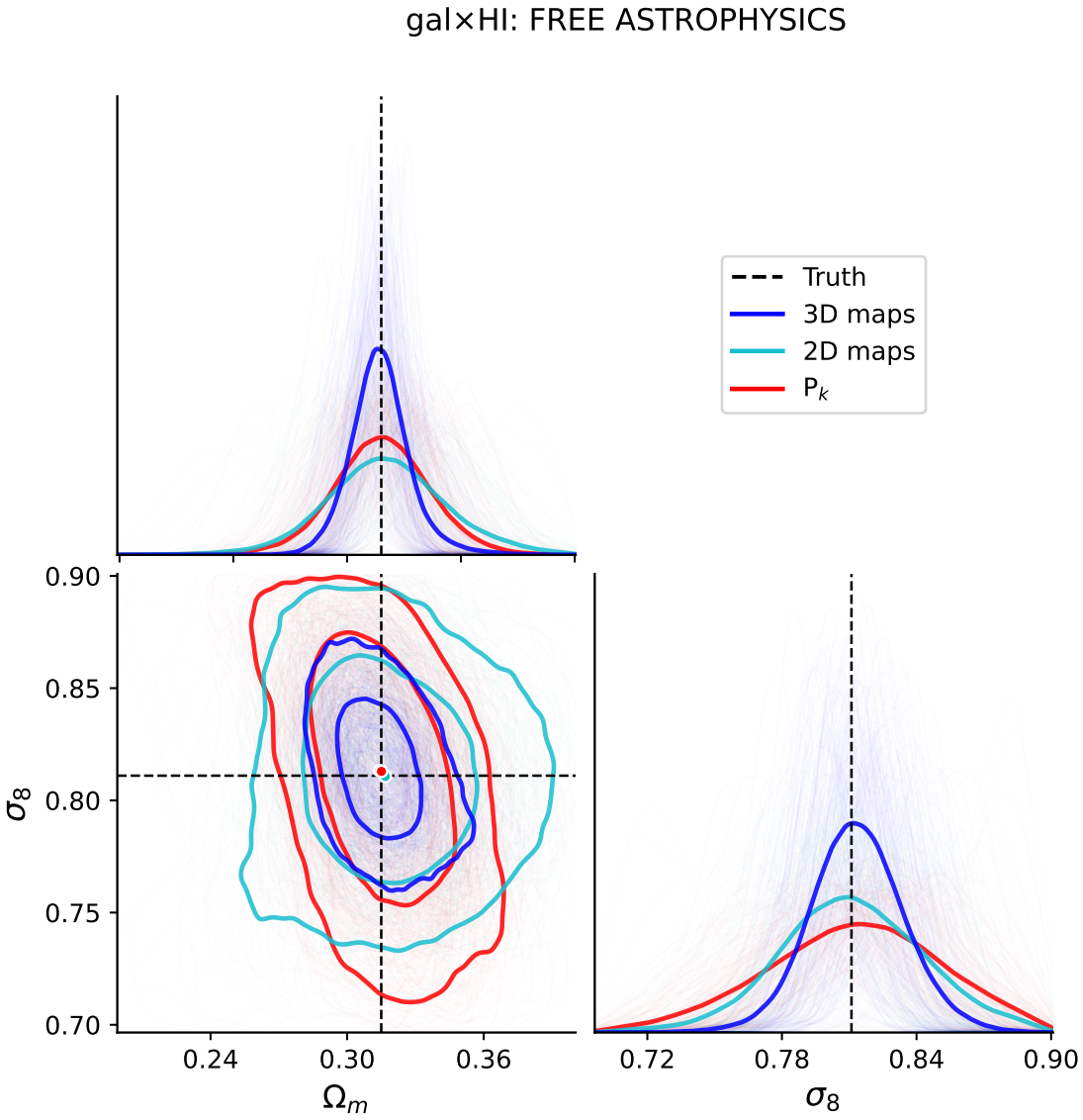


Figure 4.4: As in figure 4.3, but for the free astrophysics scenario, with astrophysical parameters marginalized over.

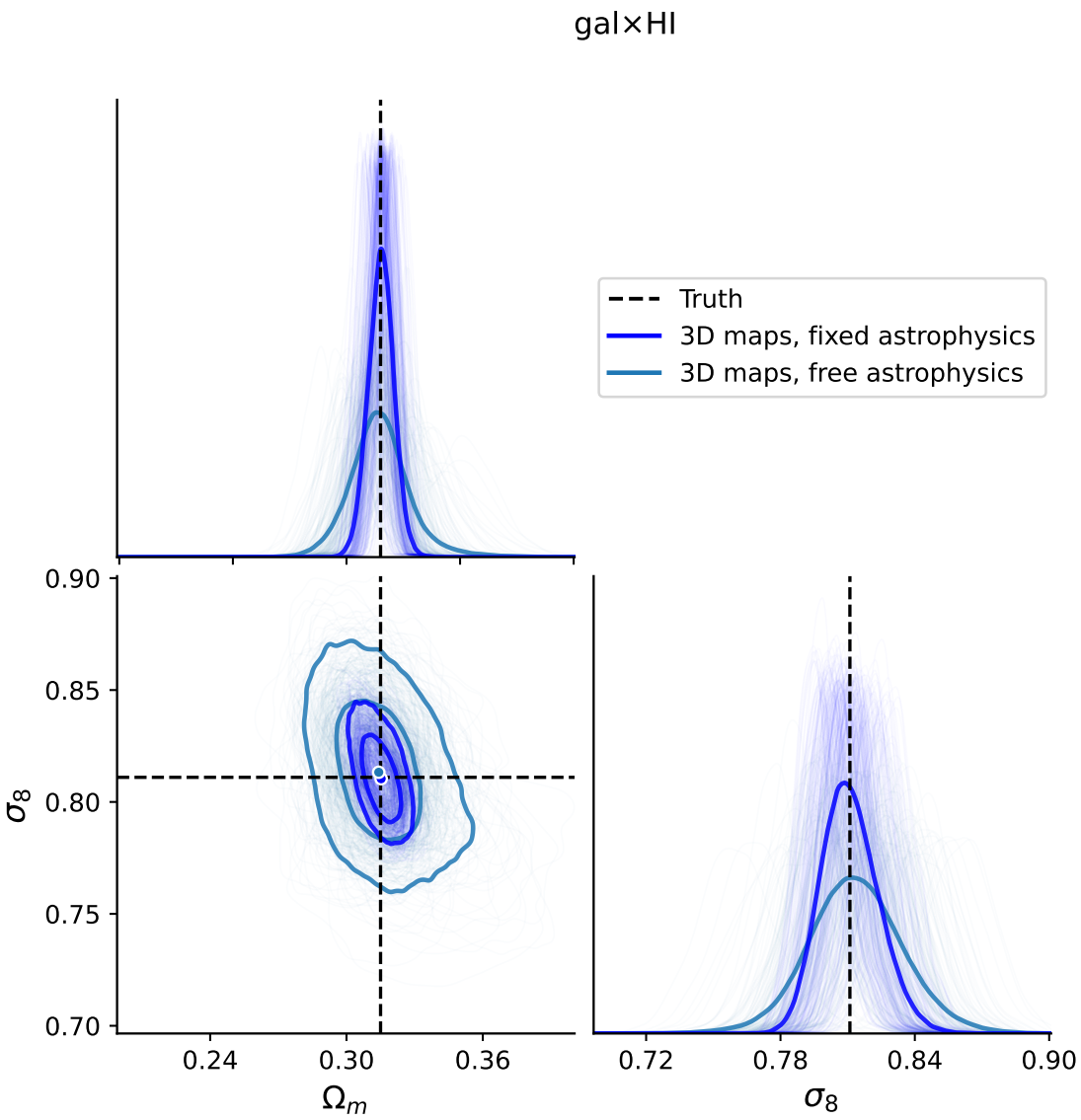


Figure 4.5: Posteriors on $\{\Omega_{\rm m}, \sigma_8\}$ marginalizing over the free astrophysics case 3D$\mathcal{M}^{\rm g\times HI}$ versus fixed astrophysics 3D$\mathcal{M}^{\rm g\times HI}_{\rm fixedA}$ for 200 validation datasets (ground truth in black). Fixing astrophysics shows smaller uncertainties as expected, but marginalized constraints retain cosmological information while being agnostic to the details of baryonic processes.

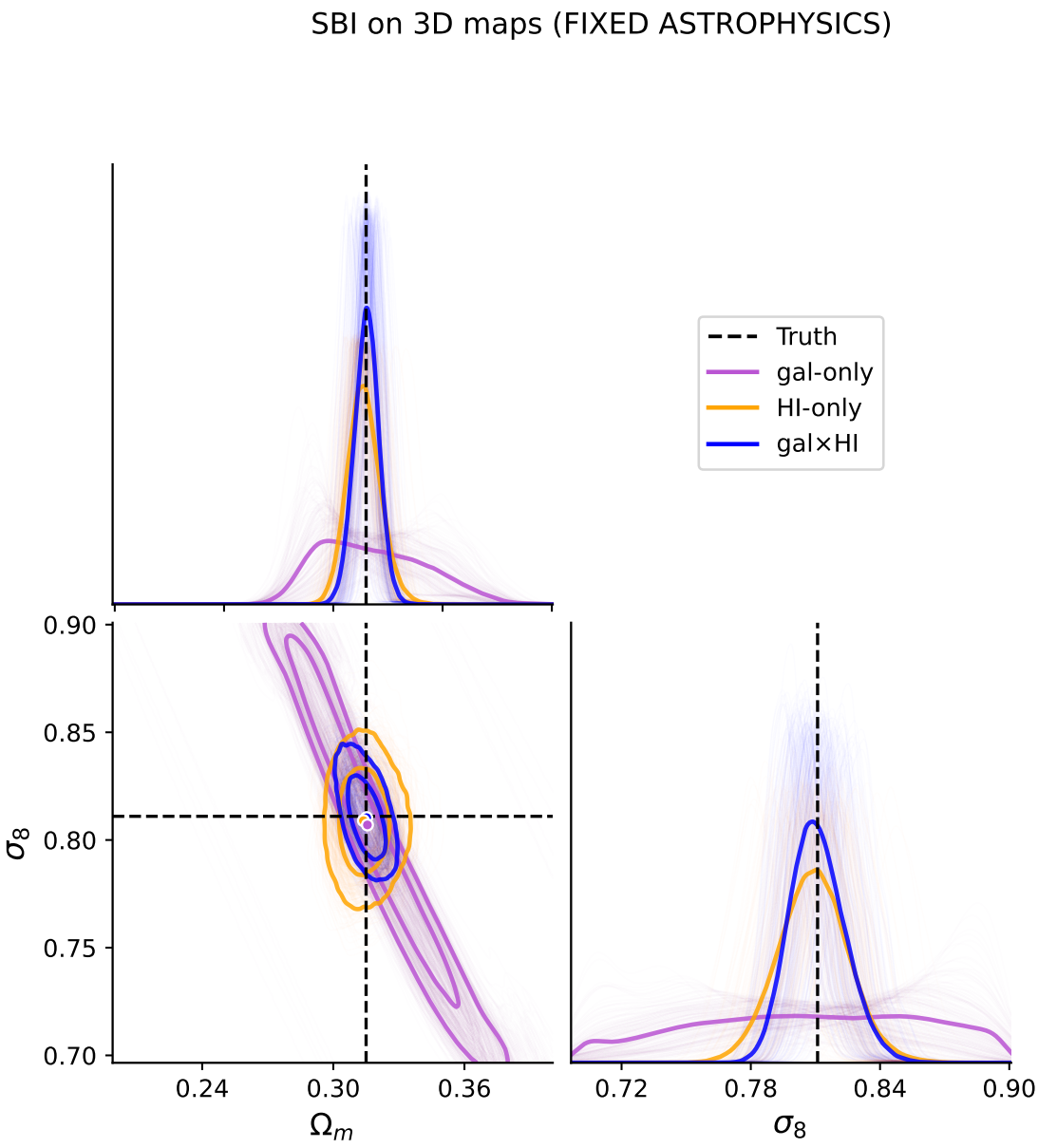


Figure 4.6: Posteriors on $\{\Omega_{\rm m}, \sigma_8\}$ (fixed astrophysics) comparing single-tracers cases $3{\rm D}\mathcal{M}^{\rm g}_{\rm fixedA}$ (purple) and $3{\rm D}\mathcal{M}^{\rm HI}_{\rm fixedA}$ (orange) with multi-tracer case $3{\rm D}\mathcal{M}^{\rm g\times HI}_{\rm fixedA}$ (blue).

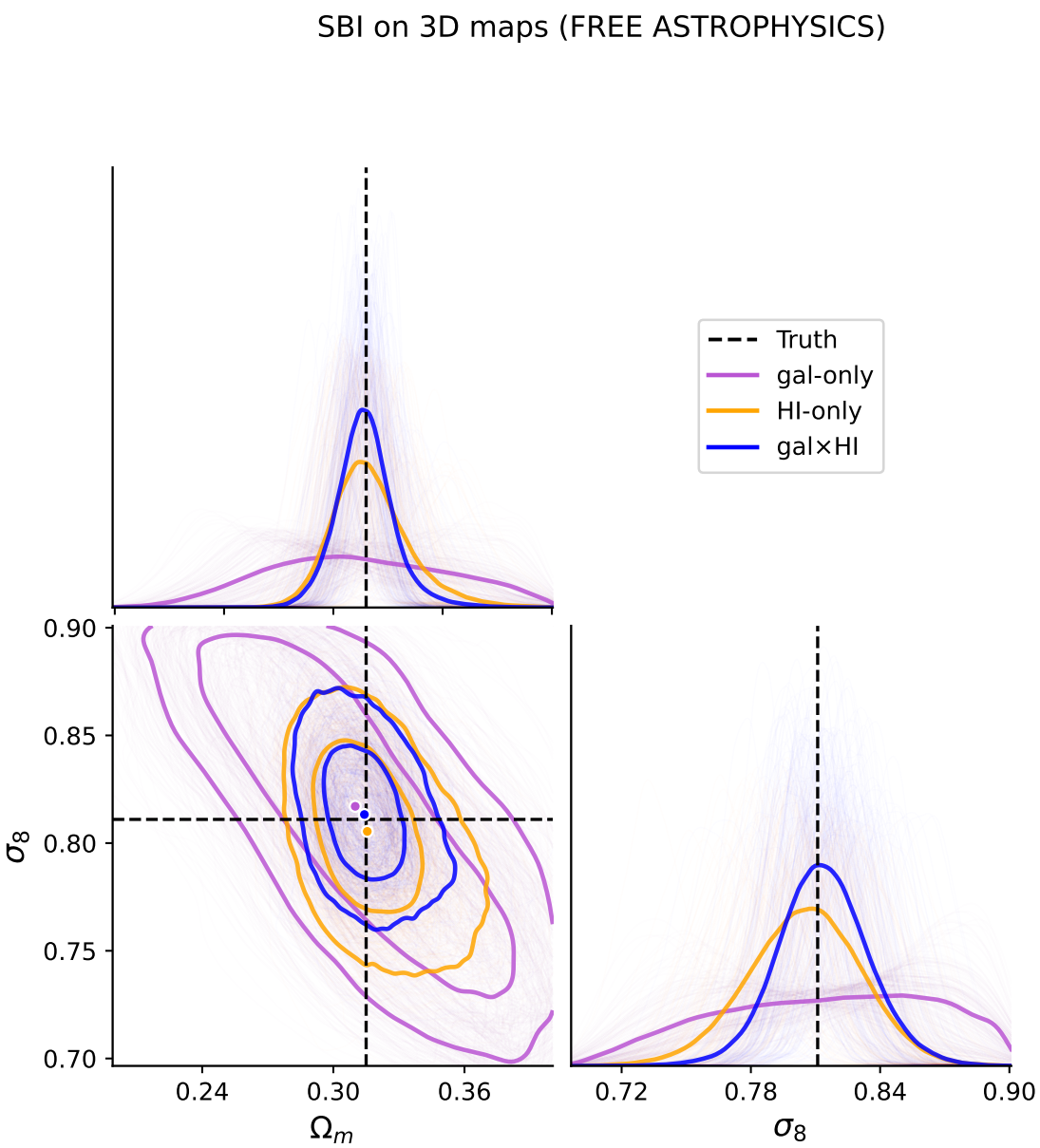


Figure 4.7: As in figure 4.6, but for the free astrophysics scenario.

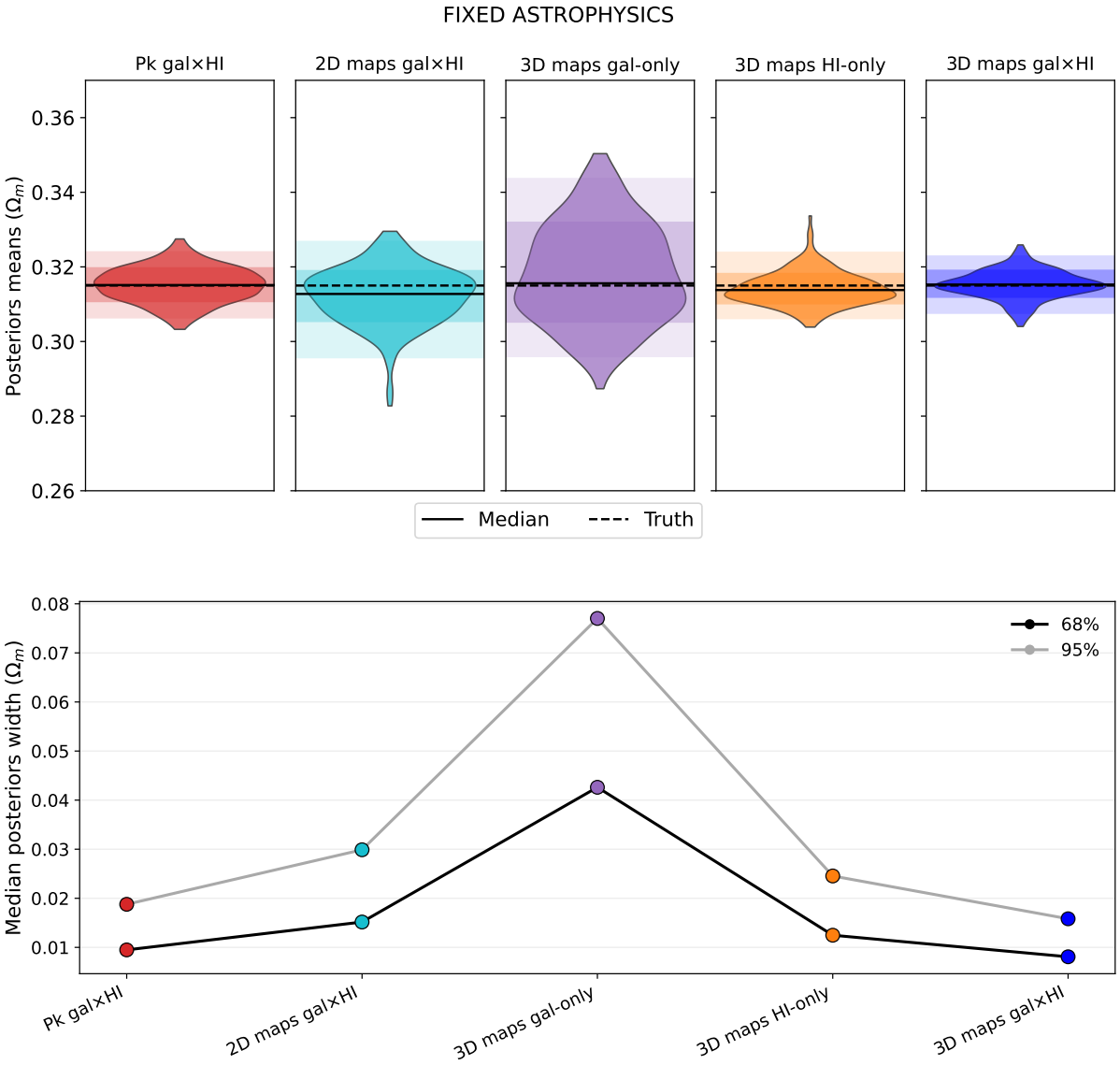


Figure 4.8: Distribution of marginal posterior summaries for $\Omega_{\rm m}$ (fixed astrophysics case). Top: violin plots showing the distribution of the posterior mean across 200 realization of the ΛCDM-Planck test set (1,000 samples for each posterior). The distance between the solid line (the median of the all posteriors samples) and the dashed line (the ground truth) can be interpreted as the bias of the inference. Bottom: median value of the marginal posteriors widths of the ΛCDM-Planck test set for 68-95% credible intervals; smaller values indicate higher precision levels.

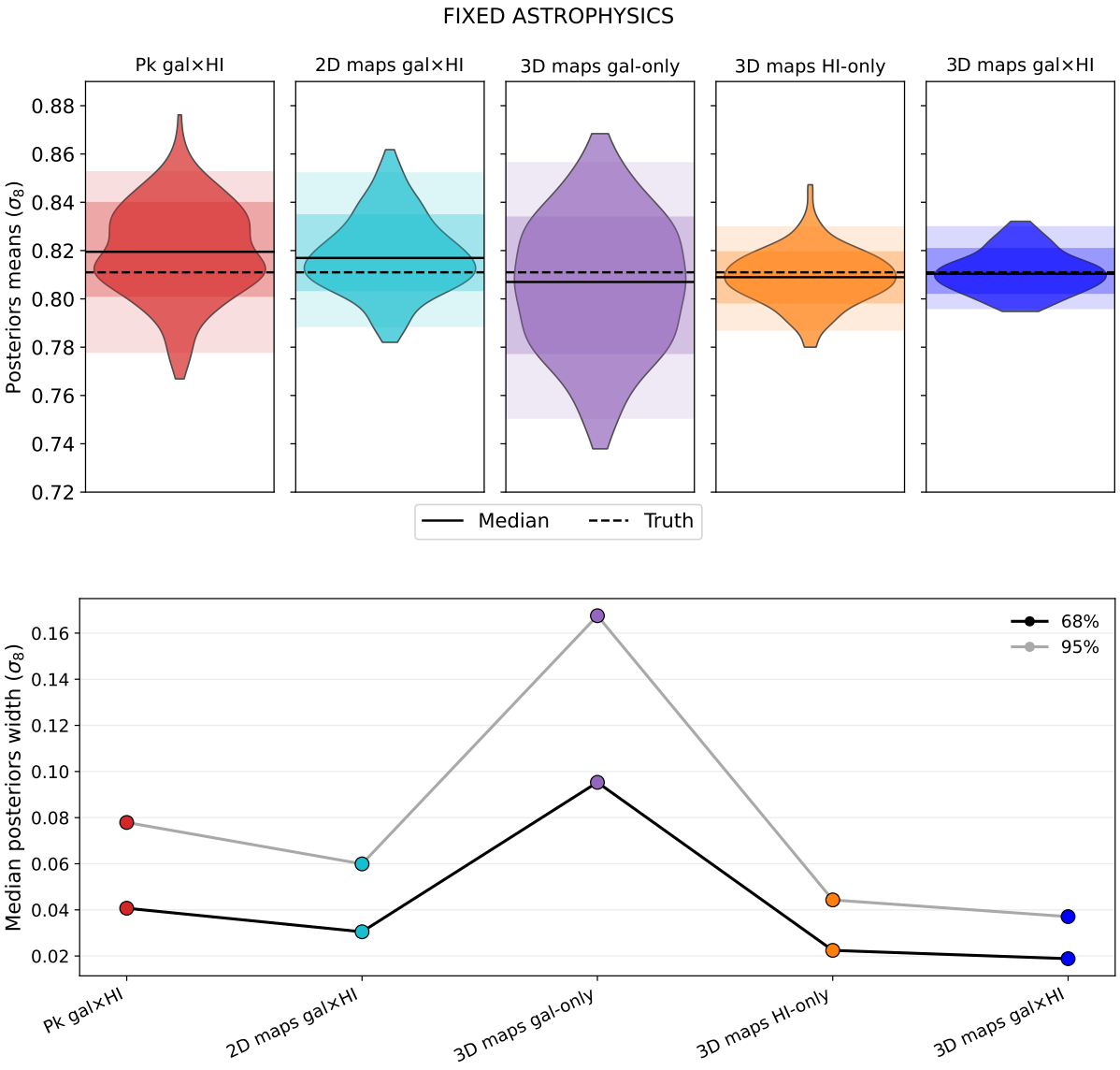


Figure 4.9: As in figure 4.8, but for σ_8.

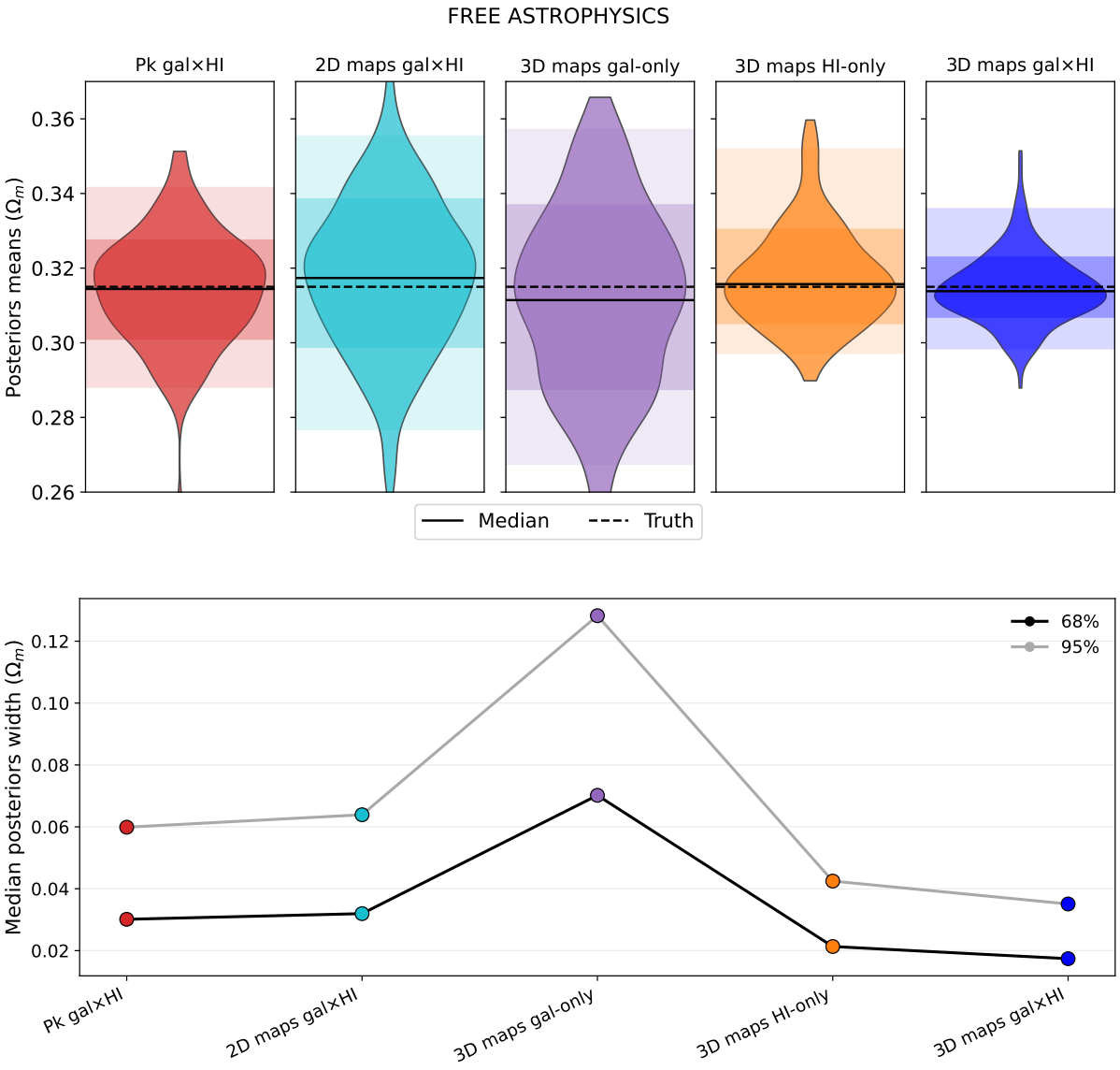


Figure 4.10: As in figure 4.8, but for the free astrophysics scenario.

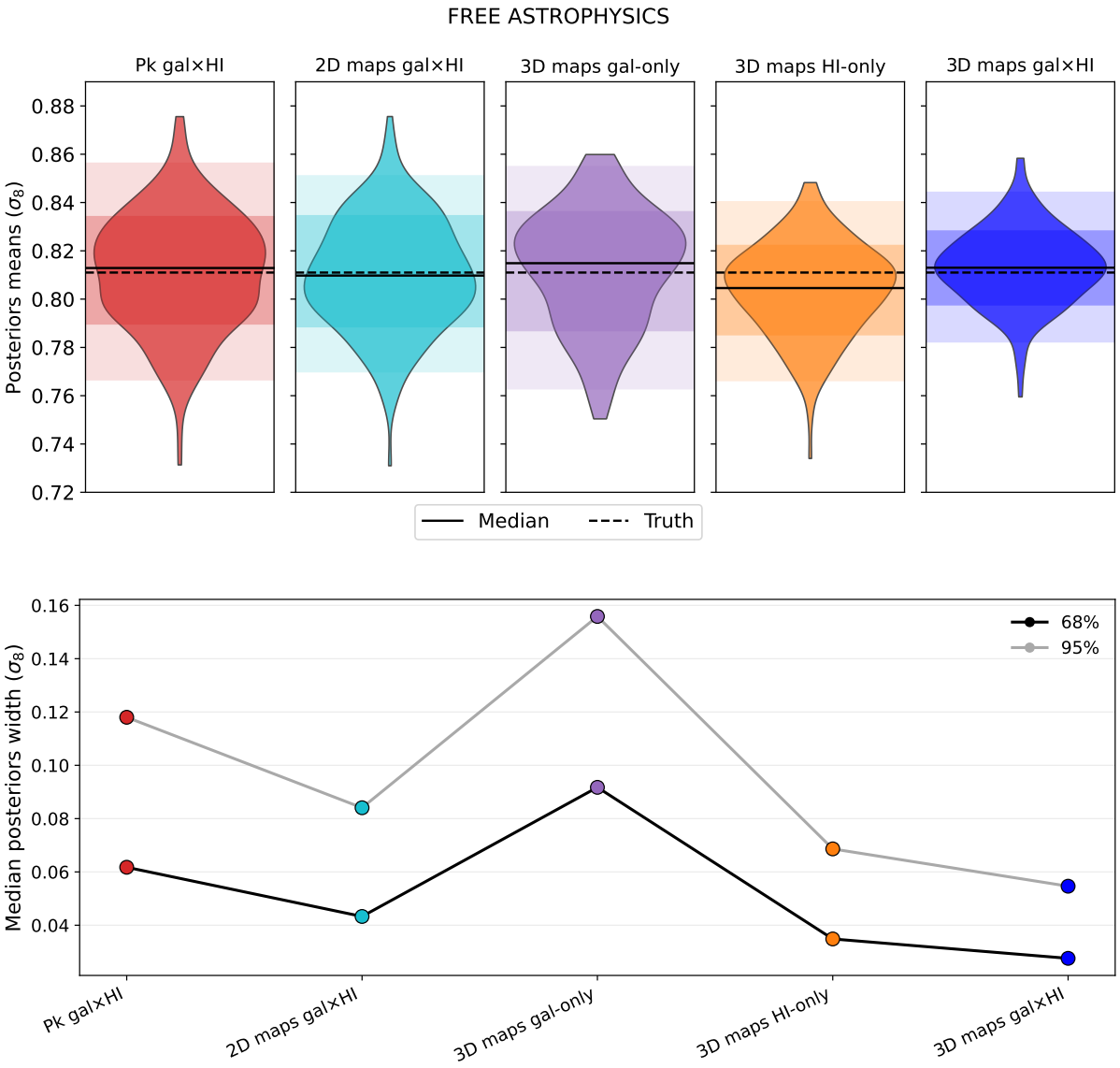


Figure 4.11: As in figure 4.10, but for σ_8.

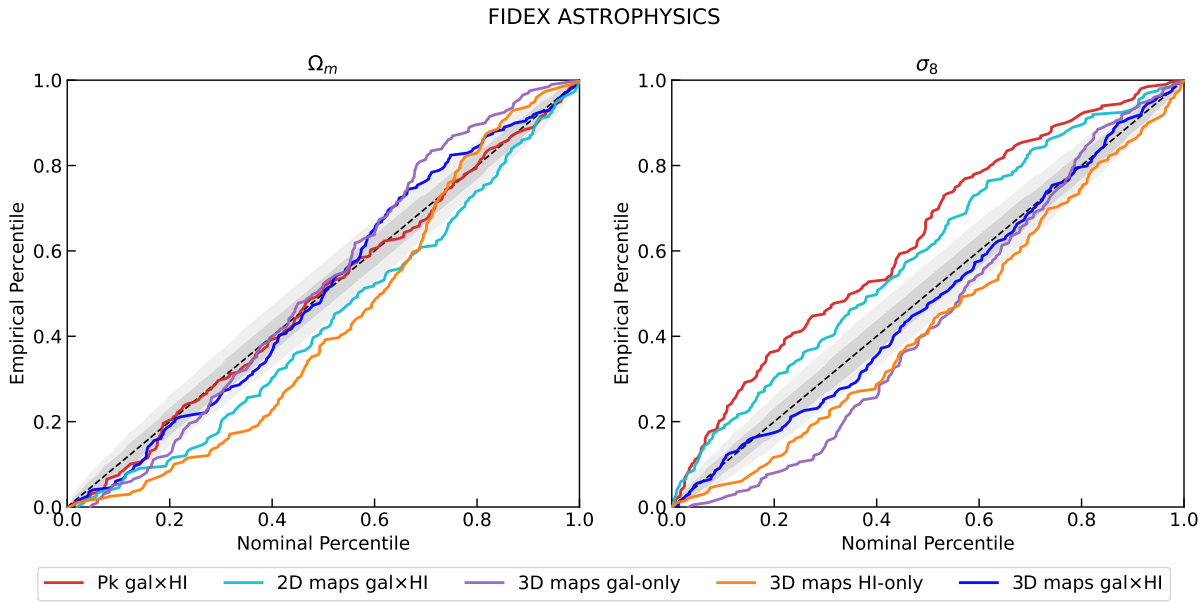


Figure 4.12: PP plots for the different data configurations for the fixed astrophysics scenario. Deviations within the gray bands are consistent with random sampling noise from a binomial distribution with 68-95% confidence level. Left panel: $\Omega_{\rm m}$. Right panel: σ_8.

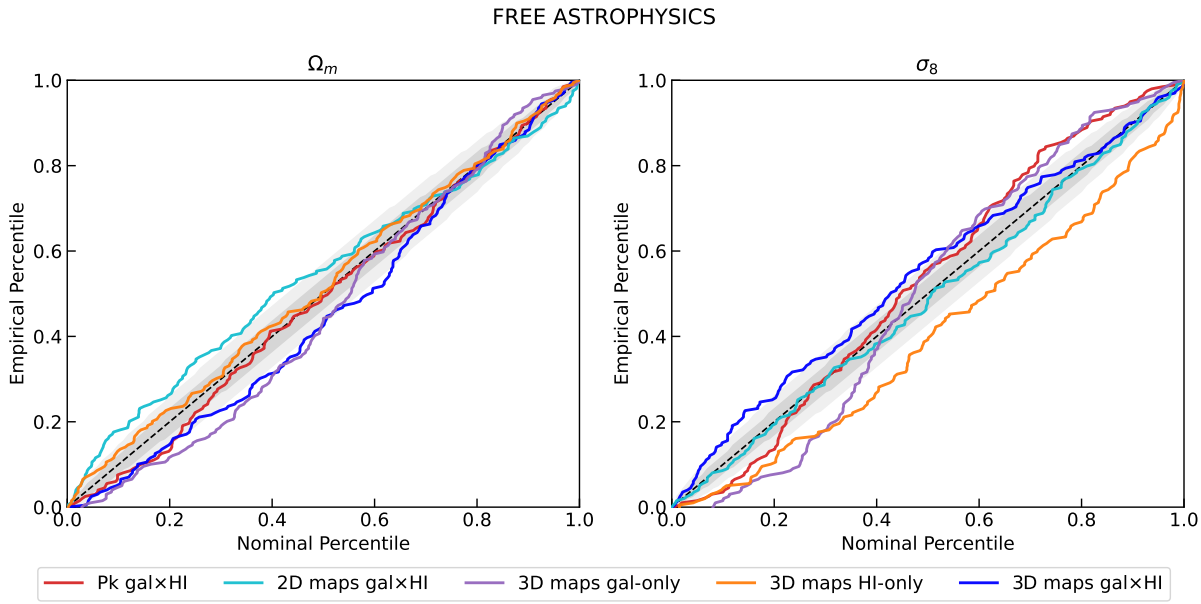


Figure 4.13: As in figure 4.12 but for the free astrophysics scenario.

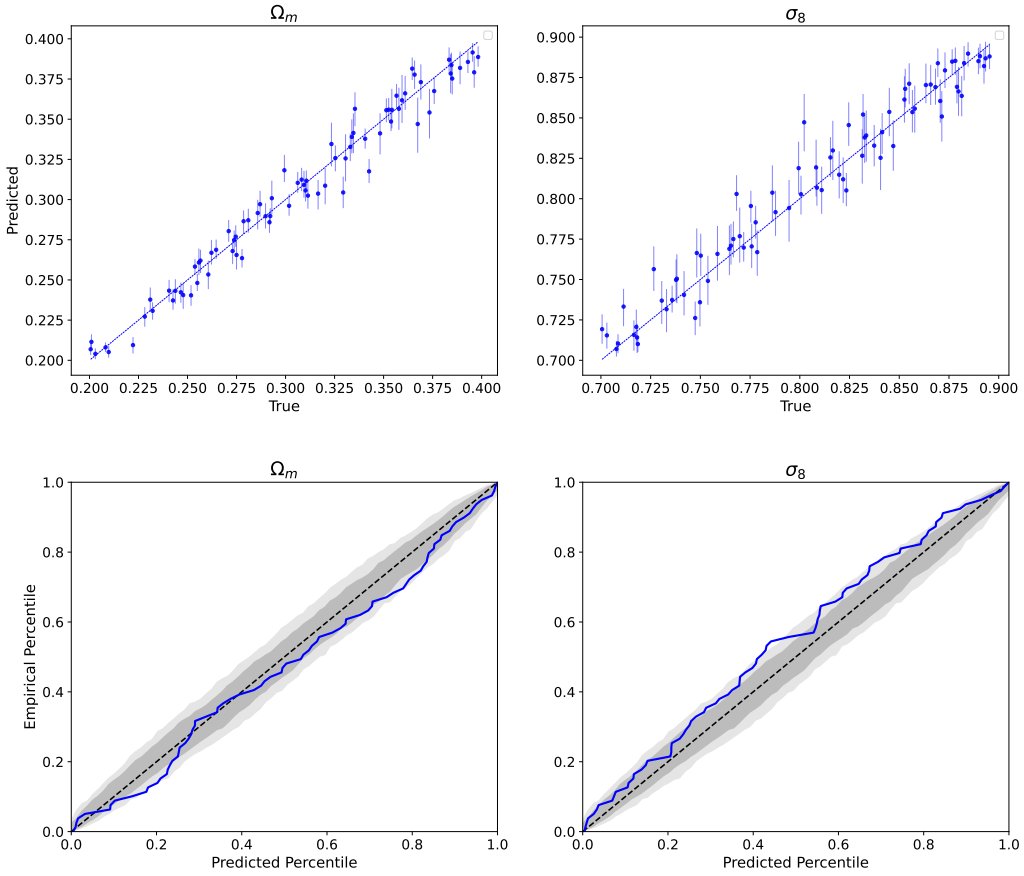


Figure 4.14: Free astrophysics case 3D$\mathcal{M}^{\mathrm{g}\times\mathrm{HI}}$, for an emulated test set of 80 maps with true parameters drawn from their priors for both cosmology and astrophysics. *Top:* each blue point shows the median of 1000 (marginal) posterior samples for a single realization, with ground truth value given by the horizontal axis. Error-bars show the 68% credible interval of the corresponding posterior. *Bottom:* PP plots with parameters drawn from their prior ranges (i.e., Bayesian PP plots, averaging over prior distributions). Deviations within the gray bands are consistent with random sampling noise from a binomial distribution with 68-95% confidence level.

4.4 Emulator

Here we describe the neural network-based technique we use as a surrogate model to replace full hydrodynamic simulations, and generate accurate 3 dimensional fields rapidly when trained on reference simulations.

4.4.1 Emulator architecture

The whole emulation pipeline is sketched in figure 4.15 and the `HGL` emulator architecture is summarized in table 4.5.

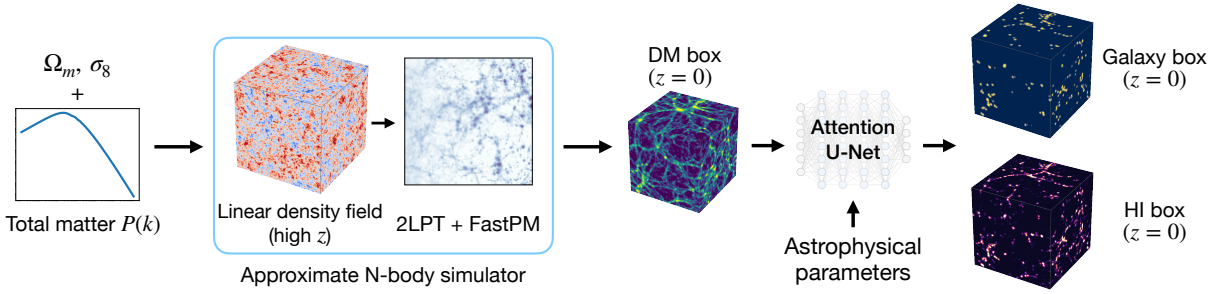


Figure 4.15: Schematics of the emulation pipeline, from the generation of fast DM maps to the `HGL` emulator. DM maps, generated by means of `FastPM`, are passed to the U-net to generate galaxies or HI, conditioning on astrophysics prescriptions.

Table 4.5: Architecture of the `HGL` emulator. All convolutions (Conv3D) are $3 \times 3 \times 3$ with circular padding. Group Normalization (GN) is used throughout, with FiLM conditioning applied within each residual block, as can be seen in the flow diagram in the last row. For stable learning, the second convolutional layer in each ResNet block is initialized to zero.

Stage	Spatial dims	Channels (in $\to$ out; groups)	Operation
Input	L^3	1	Input field
Encoder	$(L/2)^3$	$1 \to 16$ (g=1)	ResNet block + MaxPool
	$(L/4)^3$	$16 \to 32$ (g=4)	ResNet block + MaxPool
	$(L/8)^3$	$32 \to 64$ (g=4)	ResNet block + MaxPool
	$(L/8)^3$	$64 \to 128$ (g=4)	ResNet block
Bottleneck	$(L/8)^3$	$128 \to 128$ (g=4)	Self-attention
Decoder	$(L/4)^3$	$128 \to 64$ (g=4)	Upsample + ResNet block
	$(L/2)^3$	$64 \to 32$ (g=4)	Upsample + ResNet block
	L^3	$32 \to 16$ (g=4)	Upsample + ResNet block
Output	L^3	$16 \to 1$ (g=1)	ResNet block
ResNet block:	$x \to$ GN $\to$ FiLM $\to$ Conv3D $\to$ GN $\to$ FiLM $\to$ Conv3D $+ x$		

Following the network architecture of [144], the model is based on a 3D U-Net architecture comprising four encoding and four decoding residual convolutional blocks. Each encoding block consists of two 3D convolutional layers, each followed by Group Normalization and a non-linear activation, and is followed by a downsampling operation

implemented via max pooling. The decoder mirrors this structure, with each block consisting of two convolutional layers and Group Normalization layers, and spatial resolution is restored via upsampling, using trilinear interpolation.

Furthermore, to encode the cosmological and astrophysical parameters in this network, we use the FiLM (Feature-wise Linear Modulation) technique [279], which allows to condition the output of the neural network on said parameters. Let $\mathbf{x} \in \mathbb{R}^{C \times D^3}$ denote a feature layer within the residual block with C channels and D^3 dimensions, and $\mathbf{c} \in \mathbb{R}^{d_c}$ denote the vector of d_c simulation parameters. Then the FiLM transformation is defined as:

$$\mathrm{FiLM}(\mathbf{x}, \mathbf{c}) = \boldsymbol{\gamma}(\mathbf{c}) \odot \mathbf{x} + \boldsymbol{\beta}(\mathbf{c}), \tag{4.3}$$

where $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^C$ are the channel wise scaling and shift parameters of the transformation, and $\odot$ denotes element wise multiplication across channels. In each ResNet block, two of these affine transformations are applied to feature layers, with one pair of γ, β for each channel output at each convolution layer. For example, at the spatial dimension $(L/2)^3$ in the Unet, there are 16 pairs of FiLM transformations, as there are 16 channels in that ResNet block. The FiLM parameters are generated from the conditioning vector via a Multi-Layer Perceptron (MLP) f_θ:

$$[\boldsymbol{\gamma}(\mathbf{c}), \boldsymbol{\beta}(\mathbf{c})] = f_\theta(\mathbf{c}). \tag{4.4}$$

The conditioning vector $\mathbf{c}$ consists of both cosmological and astrophysical parameters, concatenated into a single vector $\mathbf{c} = [\theta_\mathcal{C},\ \theta_\mathcal{A}] \in \mathbb{R}^{d_c}$ (for our work $\mathbf{c}$ has a dimension of 6). This combined vector is passed through a shared MLP to produce the FiLM modulation parameters. This design replaces fixed affine transformations in normalization layers with conditioning-dependent transformations, allowing the network to adapt its feature representations dynamically based on the input parameters.

Given a DM map $\mathcal{F}_i^{\mathrm{DM}}(\Omega_{m_i}, \sigma_{8_i})$ characterized by a cosmology i, the emulator is able to provide the corresponding galaxy/HI map for that same cosmology i and any astrophysics j of choice:

$$\mathcal{F}_i^{\mathrm{DM}}(\Omega_{m_i}, \sigma_{8_i}) \xrightarrow{\text{U-Net}} \mathcal{F}_i^{\mathrm{gal/HI}}(\Omega_{m_i}, \sigma_{8_i}, \mathrm{A_{SN1_j}}, \mathrm{A_{AGN1_j}}, \mathrm{A_{SN2_j}}, \mathrm{A_{AGN2_j}}) \tag{4.5}$$

It is important to stress that, although the emulator is able to paint galaxies/HI for any reasonable astrophysical parameters, the cosmology between input and output must stay the same. In other words, the following emulation process is not possible:

$$\mathcal{F}_i^{\mathrm{DM}}(\Omega_{m_i}, \sigma_{8_i}) \not\rightarrow \mathcal{F}_i^{\mathrm{gal/HI}}(\Omega_{m_{k\neq i}}, \sigma_{8_{k\neq i}}, \mathrm{A_{SN1_j}}, \mathrm{A_{AGN1_j}}, \mathrm{A_{SN2_j}}, \mathrm{A_{AGN2_j}}) \tag{4.6}$$

To preserve spatial information across scales, skip connections are employed by con-

catenating feature maps from each encoder block to the corresponding decoder block at the same resolution. This design significantly improves reconstruction fidelity by retaining high frequency information lost during downsampling.

At the bottleneck, again following the previous work, we incorporate a self-attention module that enables the network to capture long range dependencies and non local correlations between voxels in the latent representation, thereby enhancing its ability to model large scale structure.

4.4.2 Emulator training

This section provides details on the emulator training. The input `FastPM` DM maps (described in section 4.2.2) of size $25\,h^{-1}$Mpc are generated for the same ICs seeds of the `CAMELS` LH suite. The output (target) galaxy/HI maps (described in section 4.2.3) of same size are instead generated from the corresponding `CAMELS` full-hydrodynamical simulations of the same suite.

The train and validation set maps derive from splitting the 508 different seeds spanning different values of the cosmological (for DM) and cosmological+astrophysical (for galaxies/HI) parameters within the `CAMELS` priors[3], in such way to have similar distributions of parameters values between train+validation and test sets. To help training, cubes in train+validation sets are repeated for all 24 possible orientations. The train and validation sets combined are then composed by $508 \times 24 = 12,192$ cubes, which are then split in a 75% and 25% train and validation split. The test set is instead composed of 242 cubes without rotations. Train, validation and test sets are composed of only 1 channel, corresponding to the tracer considered. All maps are normalized with standard scaling, fitted on each respective train set.

Following [144], the loss function used to train the emulator is a linear combination of multiple Quantile Losses (QL):

$$\mathcal{L} = \sum_{i=1}^{3} \mathrm{QL}(y, \hat{y}; \tau_i) \tag{4.7}$$

where given y and $\hat{y}$ as the target and predicted voxel value, the quantile loss is defined as:

$$\mathrm{QL}(y, \hat{y}; \tau) = \frac{1}{N} \sum_{j=1}^{N} \max\left(\tau(y_j - \hat{y}_j),\, (\tau - 1)(y_j - \hat{y}_j)\right) \tag{4.8}$$

This allows the model to learn different regions of the conditional distribution of the target field, improving robustness to sparsity. We use three quantile levels given by

[3] `CAMELS` priors for LH suite: $\Omega_\mathrm{m} \in [0.1, 0.5]$, $\sigma_8 \in [0.6, 1.0]$, $A_\mathrm{SN1} \in [0.25, 4.0]$, $A_\mathrm{AGN1} \in [0.25, 4.0]$, $A_\mathrm{SN2} \in [0.5, 2.0]$, $A_\mathrm{AGN2} \in [0.5, 2.0]$.

$\tau_i \in \{0.45, 0.60, 0.75\}$ for the prediction of HI fields and $\tau_i \in \{0.45, 0.60, 0.80\}$ for the prediction of galaxy fields. These values are chosen from empirical experience. When training the emulator, we save models checkpoint every time loss is improved. This lets us identify a-posteriori the checkpoint which corresponds to the best prediction of the power spectrum (i.e. lowest residuals between the predicted Pk and the true one) and the best checkpoint in terms of minimum of the loss (i.e. best predicted maps at pixel level). These two checkpoints may not necessary coincide, as the case of the galaxy emulator, whereas we have one single best checkpoint for the HI emulator. Generating a 64^3 voxels map takes $\sim 10^{-2}$ seconds on an A100 Nvidia GPU.

Figures 4.16 and 4.17 show the emulator performances on $25\,h^{-1}$Mpc maps for galaxies and HI respectively. The top row of each figure shows an exemplificative comparison of original `CAMELS` map from the LH suite and its corresponding emulated counterpart. The bottom row shows, on the left side, the average power spectrum residual $\Delta_{\mathrm{P_k}} = (P_k^{\mathrm{TRUE}} - P_k^{\mathrm{EMULATED}})/P_k^{\mathrm{TRUE}}$ [%] on 242 test maps taken from the LH set, with 1-σ shaded band, while on the right side the histograms of pixels values, for original and emulated test maps. Residuals are below 25% for most of the considered scales with well-centered average values, and good agreement at the histogram level.

Figures 4.18 and 4.19 are the equivalent of figures 4.16 and 4.17 but for the $50\,h^{-1}$Mpc emulated maps with the same resolution, generated by adopting the same emulator trained on the $25\,h^{-1}$Mpc ones, as motivated in section 4.2.3. Given the absence of the LH set of `CAMELS` simulations of this larger size, the test set for figures 4.18 and 4.19 is given by the CV suite (i.e. 27 seeds at same fiducial cosmology). Although the pixel-values histogram is still well calibrated, the power spectrum appears less accurately emulated than in the $25\,h^{-1}$Mpc case. This is likely due to the fact that larger modes (e.g., larger halos) present in the bigger boxes are never seen during training on the smaller box sizes. Nonetheless, given the impossibility to train an emulator on bigger boxes, as explained in section 4.2.3, we adopt this configuration for the SBI analysis. This is not problematic insofar as we are using test maps generated from the same emulator, as we do here. However, for the future this issue will need to be addressed in order to apply our framework to realistic maps.

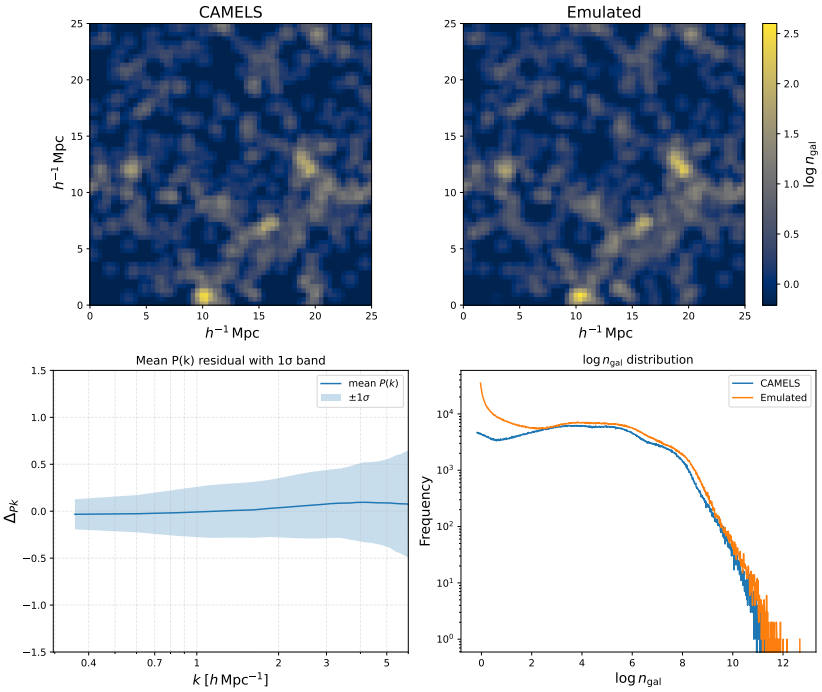


Figure 4.16: Emulator performance on CAMELS simulations for $25\,h^{-1}$Mpc box size. The top row compares the CAMELS and emulated galaxy number count fields, for an exemplificative map. The bottom left compares the power spectrum residual $\Delta_{\rm Pk}$ between CAMELS from the test set and emulated fields, and the bottom right compares the voxel-wise histogram.

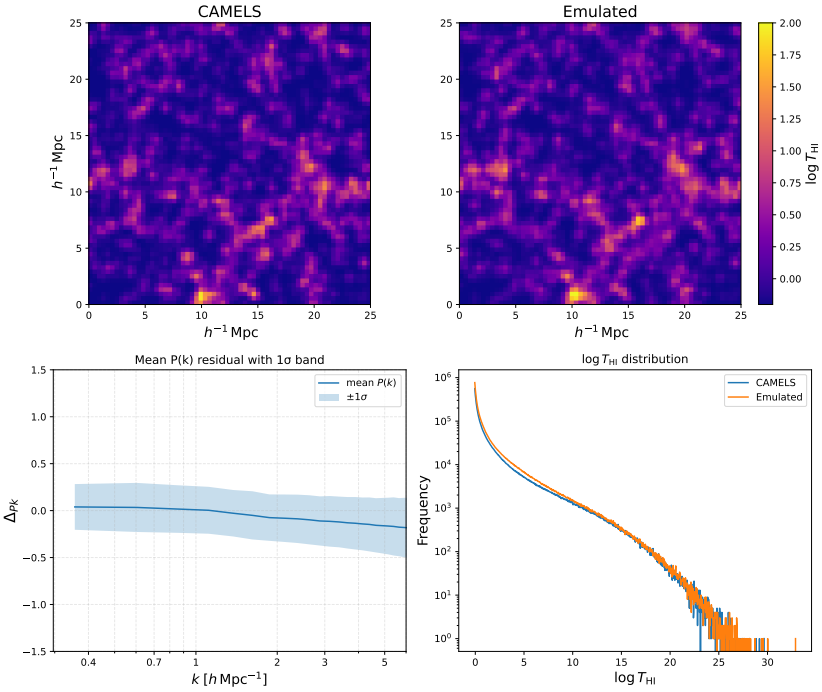


Figure 4.17: As in figure 4.16, but for the emulated HI temperature fields.

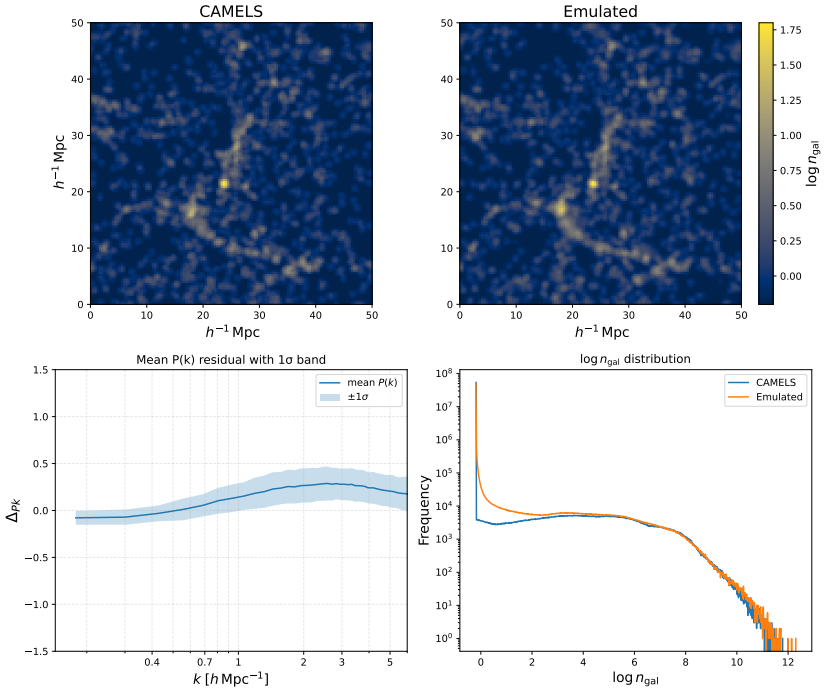


Figure 4.18: As in figure 4.16, but for $(50\,h^{-1}\mathrm{Mpc})^3$ maps, generated with the emulator trained on same resolution $(25\,h^{-1}\mathrm{Mpc})^3$ maps. Test set is the CV `CAMELS` suite.

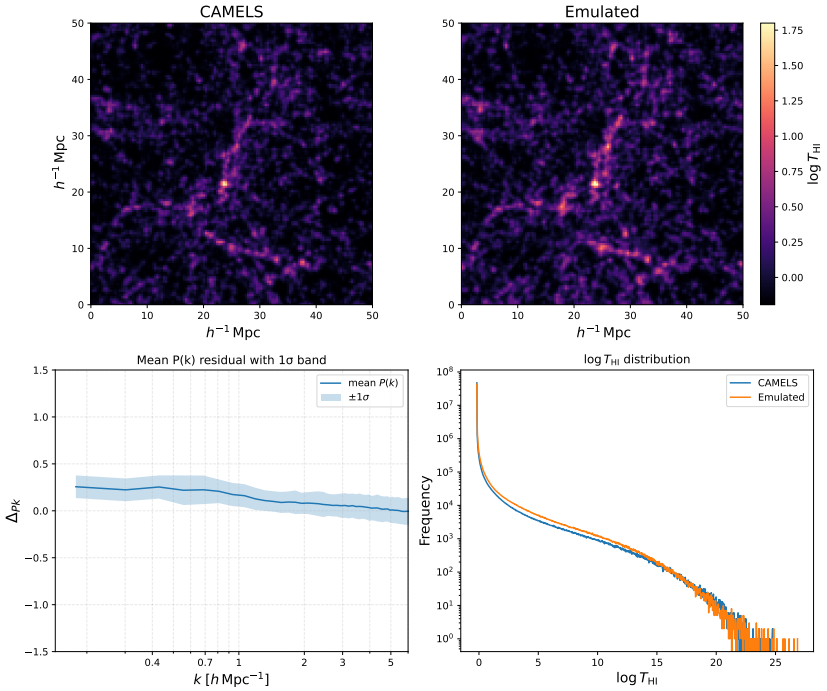


Figure 4.19: As in figure 4.17, but for $(50\,h^{-1}\mathrm{Mpc})^3$ maps, generated with the emulator trained on same resolution $(25\,h^{-1}\mathrm{Mpc})^3$ maps. Test set is the CV `CAMELS` suite.

4.5 Discussion and conclusions

We have developed an SBI pipeline combining fast dark matter simulations, neural emulators for galaxy and HI fields, and neural density estimators to perform inference on both the power spectrum and latent representations of field-level data. This setup allows us to systematically assess the impact of data compression and multi-tracer information on cosmological constraints in an idealized setting.

We find that combining tracers through cross-correlations leads to a clear improvement in constraining power. Joint galaxy–HI analyses consistently outperform single-tracer cases, with a significant reduction of parameter degeneracies. This confirms the advantage of multi-tracer approaches in extracting complementary information from the same underlying matter distribution. Moving from power spectra to field-level inference yields an additional and substantial gain. While SBI on the power spectrum provides a useful baseline, it does not capture the full information content of the data, especially when marginalizing over astrophysical effects. In contrast, SBI analysis of latent representations of the 3D fields leads to more precise and accurate posteriors, especially for σ_8. We also find that flattening the 3-dimensional structure of the field-level data into 2-dimensional projected maps does not provide the same gains, thus indicating that the improvement is linked to the three-dimensional structure of the fields. This increased constraining power comes, however, at a significantly higher computational cost.

Astrophysical uncertainties play a critical role. When marginalizing over astrophysical parameters, the constraining power of the power spectra degrades substantially, in some cases leading to uninformative posteriors. By contrast, field-level inference retains a significant fraction of the cosmological information, suggesting that non-Gaussian features help disentangle cosmology from astrophysics.

Our analysis is performed in an idealized proof-of-concept setting, neglecting observational effects such as instrumental noise, survey systematics, and foregrounds. While this allows for a clean assessment of the intrinsic information content, it also means that our current results represent an optimistic, best-case scenario. Incorporating more realistic noise modeling and moving towards survey-like simulations is necessary to assess the applicability of this approach to realistic data. For example, one would have to to apply the beam effect and thermal noise of the instrument for an SKA-like observation to HI maps (e.g. [280]), and consider the important impact of foregrounds [281]. Our approach should also be compared with existing likelihood-based pipelines [282]. For application to realistic galaxy surveys, it will be crucial to include selection functions, as well as either spectroscopic or photometric redshift errors to the mock sample. All of these improvements require a dedicated set of simulations on which to re-train our models, and we thus defer this investigation to a future work.

Finally, we emphasize that in this work we limited ourselves to a self-consistent proof-of-concept setup, where both training and inference are performed on emulated maps generated within the same framework. We did not attempt to perform a transfer learning analysis to the different simulation suites (e.g., applying SBI trained on emulated maps to original `CAMELS` maps or to other suites), since previous studies [217] have highlighted the sensitivity of field-level SBI methods to out-of-distribution effects and small-scale discrepancies between simulations, even just for pure N-body codes. This mean that transfer learning between emulated and simulated maps (and even more, real observations) remains challenging in the absence of dedicated mitigation strategies. Addressing these important issue requires a separate investigation, beyond the scope of the present work, whose goal was instead to establish and validate the proposed field-level multi-tracer SBI framework within the context of a given emulated suited of simulations.

Overall, our results highlight the combined impact of multi-tracer analyses and field-level SBI in unlocking the information content of large-scale structure. Extending these methods to more realistic settings and improving their computational efficiency will be key for their application to upcoming cosmological surveys.

Chapter 5

Conclusions

Field-level inference of LSS is a powerful and promising tool for cosmology; nonetheless, it requires forward models that are simultaneously accurate, cheap enough to be evaluated many thousands of times, and reliable on a range of the parameters being inferred rather than at one fixed cosmology. Full-hydrodynamical simulations satisfy the first and third requirements but drastically fail on the second; approximate solvers satisfy the second and the third but fail on the first. This thesis has been concerned with developing a full pipeline for field-level inference, aiming at bridging the gap between full-hydrodynamical and approximate simulations and then implementing an SBI pipeline for robust field-level inference.

The first ingredient is developed in Chapter 2, which discusses `JERALD`, an effective displacement model that corrects the output of a fast approximate solver towards a full-hydrodynamical reference. Crucially, the model only uses at most tens of physically-interpretable parameters, but can reproduce dark matter, stellar mass and neutral hydrogen fields from higher-resolution full-hydrodynamical simulations down to very non-linear scales. Chapter 4 provides the second ingredient, building a field-level multi-tracer SBI pipeline using approximate dark matter maps and a neural emulator for the observable fields, and finds that preserving the full three-dimensional structure of the field yields substantially more information than any two-point compression, that combining galaxy and HI tracers tightens the constraints and alleviates their bias, and that the advantage of field-level inference is largest when marginalizing over astrophysical uncertainties.

Chapter 3 attempts to close the loop by making `JERALD` depend on cosmology, but with no success. Nonetheless, the analysis establishes what its loss constrains, i.e. a combination of the model parameters, $J(k)$, together with γ, rather than all parameters individually, and quantifies how much of $J(k)$ and γ can be predicted from the properties of a simulation it is trained on and from its cosmological and astrophysical parameters. Two results from this chapter have a direct effect on the pipeline of Chapter 4: $J(k)$ carries significantly more information about the cosmological and astrophysical parameters than

the individual model parameters do, and a single fixed model parameter set transfers well at the level of ensemble means and two-point statistics, and considerably less well map by map.

The natural way to unify the two lines is therefore to insert `JERALD` at the first stage of the pipeline of Chapter 4, where approximate dark matter maps are produced before the emulator sees them. As currently constructed, that pipeline carries the entire dependence on baryonic physics inside the learned weights of the neural emulator; the analysis of Chapter 3 indicates that `JERALD` can express part of this dependence through a few interpretable quantities: the amplitude of galactic wind feedback retains a dependence on one of the model parameters, γ, that survives conditioning on every property of the input field, precisely because it modifies the reference and not the approximate simulation, and the cosmological and astrophysical parameters contribute measurably to $J(k)$. Supplying the emulator with a dark matter field that has already been corrected both for the coarse-graining effects of the approximate N-body simulation and for baryonic feedback in a physically motivated and parameter-dependent way would lift part of the numerical and astrophysical corrections from the network and move it into a model whose behavior can be inspected.

There is also a more specific argument for this integration. The pipeline of Chapter 4 was validated self-consistently, with training and inference both performed on emulated maps, and the transfer of a field-level inference model to the underlying full-hydrodynamical simulations was not attempted. The emulator comparisons of Section 4.4.2 illustrate the problem: agreement with the reference is good in the mean, and less good for individual realizations. Since inference at the field level acts on single maps rather than on ensemble averages, per-realization fidelity is then a fundamental requirement, and it is exactly what an effective displacement model fitted to reproduce a particular realization is designed to provide. The obstacle to doing this at the time of writing this thesis is the one Chapter 3 describes. Applying `JERALD` with a fixed parameter set would improve the mean behavior of the input maps but would inherit the realization-to-realization scatter documented there, which is exactly the quantity the integration is meant to reduce; applying it with parameters optimized *per simulation* would defeat the purpose, since the aim of an emulator is precisely to avoid running full-hydrodynamical simulations for each data point used by an SBI pipeline. Closing that gap for `JERALD` requires the dependence of $J(k)$ and γ on the cosmological and astrophysical parameters to be characterized well enough to be evaluated rather than fitted, and the analysis of Chapter 3 is specific about what steps are needed to reach this goal: simulation suites designed to separate cosmological signal from realization noise, a regressor with access to the phase information that power-spectrum summaries discard, and a test of whether the prediction of γ and $J(k)$ can compose into a single one without their errors compounding. None of these steps

requires a new model, and none is out of reach.

A central conclusion of this thesis, then, is that what precludes making a physically motivated correction depend on cosmology is not the correction itself, whose behavior is by now well characterized, but rather the conditions under which that dependence can be determined at all. Field-level inference will require forward models that are cheap, accurate on non-linear scales, and transparent enough to be trusted where the training data run out; whether `JERALD` provides one remains an open question, but the analysis presented here specifies what is needed to answer it.

A Cloud-In-Cell paint and readout

A.1 CIC calculation of density

Given a set of P particles with positions $\boldsymbol{r}_\mu = (x_\mu,\, y_\mu,\, z_\mu)$ and masses m_μ, $\mu = 0,\, ...,\, P-1$ and a mesh with size $(N_x,\, N_y,\, N_z)$ and box size $(L_x,\, L_y,\, L_z)$, the Cloud-in-Cell algorithm computes the contribution of particle μ to the mass of the 8 neighboring mesh vertices by distributing its mass among them based on the their distance (i.e. via trilinear interpolation):

$$\begin{aligned}
M_{i,\,j,\,k} &= (1-d_x^\mu)(1-d_y^\mu)(1-d_z^\mu)m_\mu \\
M_{i+1,\,j,\,k} &= d_x^\mu(1-d_y^\mu)(1-d_z^\mu)m_\mu \\
M_{i,\,j+1,\,k} &= (1-d_x^\mu)d_y^\mu(1-d_z^\mu)m_\mu \\
M_{i,\,j,\,k+1} &= (1-d_x^\mu)(1-d_y^\mu)d_z^\mu m_\mu \\
M_{i+1,\,j+1,\,k} &= d_x^\mu d_y^\mu(1-d_z^\mu)m_\mu \\
M_{i+1,\,j,\,k+1} &= d_x^\mu(1-d_y^\mu)d_z^\mu m_\mu \\
M_{i,\,j+1,\,k+1} &= (1-d_x^\mu)d_y^\mu d_z^\mu m_\mu \\
M_{i+1,\,j+1,\,k+1} &= d_x^\mu d_y^\mu d_z^\mu m_\mu \quad ,
\end{aligned} \tag{A.1}$$

where $(i,\, j,\, k)$ is the closest mesh point to particle μ:

$$i,\, j,\, k = \left\lfloor x_\mu \frac{N_x}{L_x} \right\rfloor,\, \left\lfloor y_\mu \frac{N_y}{L_y} \right\rfloor,\, \left\lfloor z_\mu \frac{N_z}{L_z} \right\rfloor \quad , \tag{A.2}$$

and d_a^μ, $a = x,\, y,\, z$ is the distance in mesh units to the closest mesh vertex, that is, e.g.

$$d_x^\mu = x_\mu \frac{N_x}{L_x} - i \quad . \tag{A.3}$$

Note that, the closer the particle (i.e. the smaller d_a), the stronger the contribution to the mesh point mass (with $1 - d_a$). See Figures A.1 and A.2 for visual representations. For multiple particles, each contribution is simply added linearly. Notice that the strength of each contribution does not depend on the Euclidean distance between the particle and the mesh point; this scheme guarantees that the sum of the mass deposited over the entire mesh equals the total mass of the particles.

A.2 CIC readout

The CIC readout of a 3D field $\boldsymbol{M}$ assigns a weight to a particle in position $\boldsymbol{r} = (x,\, y,\, z)$ by interpolating the values of $\boldsymbol{M}$ of its 8 neighboring mesh points in a way that is equivalent

to the CIC paint:

$$\begin{aligned} m(\boldsymbol{r}) = {} & (1-d_x)(1-d_y)(1-d_z)M_{i,\,j,\,k} + \\ & + d_x(1-d_y)(1-d_z)M_{i+1,\,j,\,k} + \\ & + (1-d_x)d_y(1-d_z)M_{i,\,j+1,\,k} + \\ & + (1-d_x)(1-d_y)d_zM_{i,\,j,\,k+1} + \\ & + d_xd_y(1-d_z)M_{i+1,\,j+1,\,k} + \\ & + d_x(1-d_y)d_zM_{i+1,\,j,\,k+1} + \\ & + (1-d_x)d_yd_zM_{i,\,j+1,\,k+1} + \\ & + d_xd_yd_zM_{i+1,\,j+1,\,k+1} \quad , \end{aligned} \tag{A.4}$$

where i, j, k are defined again as in Eq. (A.2) and d_a, $a = x$, y, z is the distance in mesh units of the particle to mesh vertex i, j, k, see Eq. (A.3).

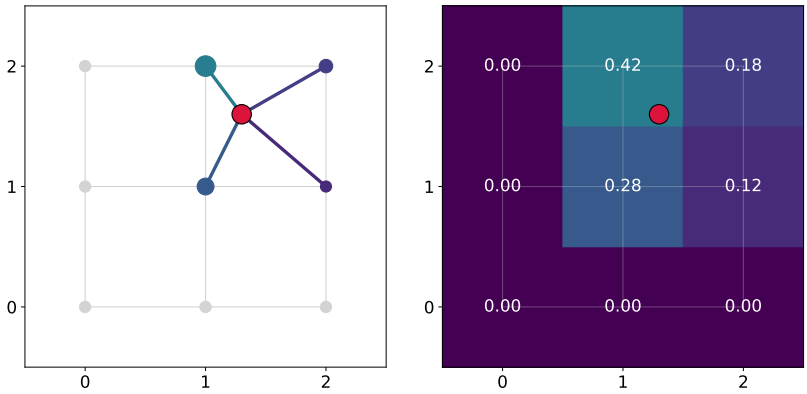


Figure A.1: Visual representation of the CIC paint algorithm in 2D. The red particle (with unit mass) contributes to its 4 neighboring mesh vertices based on its distance from them.

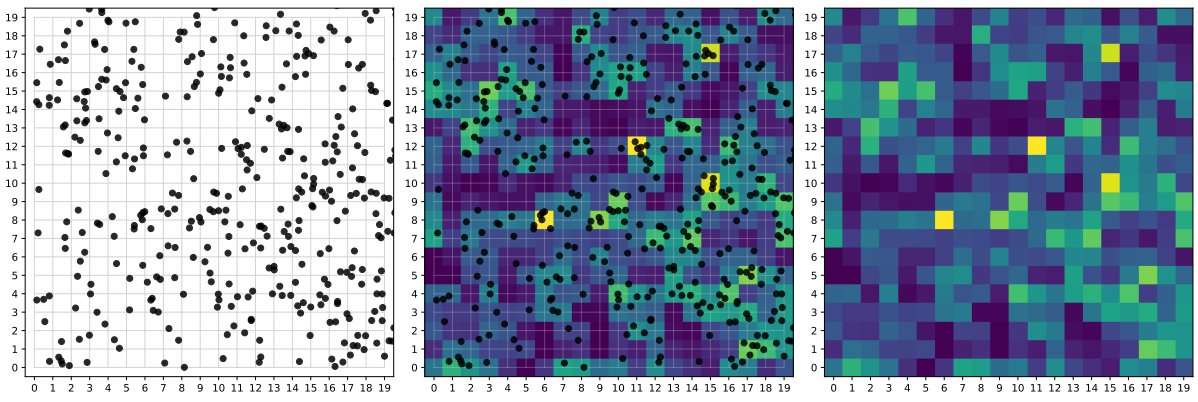


Figure A.2: Visual representation of the CIC paint algorithm in 2D for a set of particles.

B Domain decomposition and halo exchange

For the purpose of speed, especially on large clusters, the `JERALD` code is parallelized using MPI: both the particles and the mesh are distributed among the MPI ranks.

The mesh is split following the prescription used by FFTW [156], which performs a slab decomposition along the first axis, dividing the mesh into slices of width (approximately) N_x/C where N_x is the mesh size along the first axis and C is the size of the MPI communicator. Therefore, for a simulation box of size (L_x, L_y, L_z), rank m only handles the simulation box slice with edges (approximately) ($L_x/C \times m$, $L_x/C \times (m+1)$) with $m = 0, ..., C-1$ along x, (0, L_y) along y and (0, L_z) along z.

In the simplest case, the P particles are split among the ranks such that rank m only handles particles with indices from $\lfloor P/C \times m \rfloor$ to $\lfloor P/C \times (m+1) \rfloor$, with $m = 0, ..., C-1$. This splitting, however, may be inefficient, as it does not consider where the particles lie in the simulation box; for this reason, a particle may be assigned to a rank that does not handle the local part of the field containing its coordinates. Therefore, a different particle-assignment prescription was also implemented, distributing them approximately equally between the ranks, though in each rank prioritizing the ones with coordinates within the local edges of the field.

When performing a CIC paint, particles that belong to, say, rank m but contribute to the mass of adjacent mesh points living in ranks $n \neq m$ need to be copied over to these ranks (see, for instance, the green particle in Figure B.1). This operation is called *halo exchange*, and the temporary copies of particles in different ranks are called *ghosts*. To perform a halo exchange, each rank first loops over its local particles and establishes which ranks their masses contribute to. This allows to create a *decomposition layout*, i.e. a structure that contains the information needed not only to send the local ghosts to the ranks they contribute to but also to gather contributions from the parts of the field that lie in different ranks to the particles belonging to the current one for readout operations. After this information is exchanged among ranks, each rank locally paints its part of the field, also including the ghost particles received from neighboring ranks. The procedure is similar for CIC readout: after having determined which particles among all ranks receive contributions from the part of the field in the current rank, which is achieved by producing the same decomposition layout mentioned above, a local readout onto these particles is performed. Finally, all contributions for any particle from the local field are re-distributed back to the rank where the particle belongs.

Since CIC readouts are performed only after CIC paints onto the same particles' positions, the code stores the decomposition layout computed for the CIC paint and recycles it for the CIC readout, speeding up the latter.

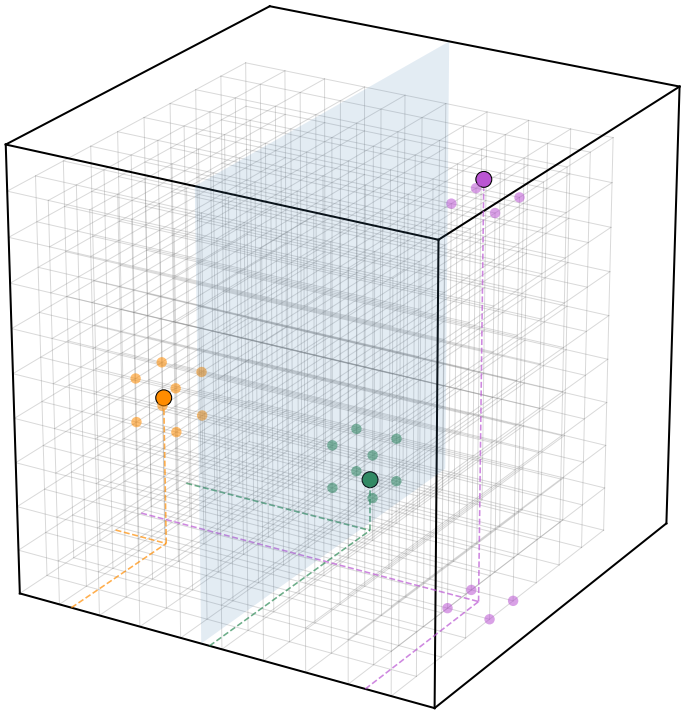


Figure B.1: Visualization of the CIC algorithm in the case of a mesh (in light gray) split between two MPI ranks. The light blue plane represents the splitting: mesh points to the left of it belong to rank, say, 1, while mesh points to the right belong to rank 2. The orange particle lies fully within the left slice, thus it does not need to be duplicated: in case it belonged to rank 2, it would only need to be sent to rank 1. The same is true for the violet particle; in this case, the visualization highlights which mesh points the mass of the particle contributes to in case one of its coordinate exceeds the mesh limits and Periodic Boundary Conditions (PBC) apply (in this case, for example, $N_z = 10$ and $L_z = 10$, thus the z coordinates of the mesh points are 0, ..., 9, and the z coordinate of the particle is $x_z = 9.4$). On the contrary, the green particle contributes to mesh points both to the left and to the right of the dividing plane, therefore it needs to be duplicated and distributed to both rank 1 and 2.

C Automatic differentiation

On the Python end, the code is written in `JAX` [155], which allows for automatic differentiation (AD). Since the code defines custom functions—specifically, parallelized CIC paint and readout, parallelized Discrete Fourier Transforms (DFTs) and parallelized power spectrum—, these need to be equipped with definitions of Vector-Jacobian Products (VJPs), which are central for reverse-mode AD. For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ with Jacobian $\boldsymbol{J}_f \in \mathbb{R}^{m\times n}$, the VJP with respect to a cotangent vector $\boldsymbol{v} \in \mathbb{R}^m$ is defined as $\boldsymbol{v}^\top \boldsymbol{J}_f \in \mathbb{R}^n$, i.e. the pullback of $\boldsymbol{v}$ through f. This construction is what makes reverse-mode AD efficient for scalar-valued loss functions: rather than materializing the full (and potentially enormous) Jacobian, the chain rule is applied backward through the computational graph, with each primitive operation contributing its own local VJP that is composed with the incoming cotangent from downstream. Concretely, for a composition $f = f_L \circ \cdots \circ f_1$, the gradient of the final scalar output is obtained by propagating $\boldsymbol{v}$ back through each f_i in reverse order, at each step computing $\boldsymbol{x}_i = \mathrm{vjp}_{f_i}(\boldsymbol{v}_i)$, so that the total cost scales with the cost of evaluating f once.

C.1 VJP of CIC paint and readout

CIC paint VJP The VJP of the CIC paint defined in Appendix A needs to be computed with respect to the input positions $\boldsymbol{r}_\mu$ and masses m_μ. Considering the positions first (only the x component—the others dimensions are equivalent):

$$\mathrm{vjp}_\mu = v^{i,j,k}\frac{\partial M_{i,j,k}}{\partial x_\mu} \quad , \tag{C.1}$$

where $M_{i,j,k}$ is the output 3D field (i, j, k are generic mesh indices at this point) and repeated indices are summed over. Clearly, only the mesh points which particle μ contributes to yield a non-zero partial derivative; the product with $\boldsymbol{v}$ then results in a sum over the 8 mesh indices which particle μ contributes to:

$$\begin{aligned} v^{i,j,k}\frac{\partial M_{i,j,k}}{\partial x_\mu} &= v^{\hat{\imath},\hat{\jmath},\hat{k}}\frac{\partial}{\partial x_\mu}(1-d_x^\mu)(1-d_y^\mu)(1-d_z^\mu)m_\mu + \cdots + v^{\hat{\imath}+1,\hat{\jmath}+1,\hat{k}+1}\frac{\partial}{\partial x_\mu}d_x^\mu d_y^\mu d_z^\mu m_\mu \\ &= v^{\hat{\imath},\hat{\jmath},\hat{k}}\left(-\frac{N_x}{L_x}\right)(1-d_y^\mu)(1-d_z^\mu)m_\mu + \cdots + v^{\hat{\imath}+1,\hat{\jmath}+1,\hat{k}+1}\frac{N_x}{L_x}d_y^\mu d_z^\mu m_\mu \quad , \end{aligned} \tag{C.2}$$

where $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$ are now the ones defined in Eq. (A.2) and again N_x and L_x are the mesh size and the box size along x: practically, this equation represents a CIC readout of the field $\boldsymbol{v}$ at position $\boldsymbol{r}_\mu$, multiplied by m_μ and replacing d_x^μ with N_x/L_x and $1-d_x^\mu$ with $-N_x/L_x$.

The derivative with respect to the mass will yield the same result, although without the substitution of d and without m_μ, thus corresponding to a plain readout of $\boldsymbol{v}$ at position $\boldsymbol{r}_\mu$.

CIC readout VJP Consider the positions first:

$$\mathrm{vjp}_\mu = v^\nu \frac{\partial m_\nu}{\partial x_\mu} \quad , \tag{C.3}$$

where m_ν is the output mass at position $\boldsymbol{r}_\nu$. Here, or course, only the term $\nu = \mu$ is non-zero, and, from Eq. (A.4), computing the derivative is simple (using the same definition for i, j, k):

$$\begin{aligned} v^\nu \frac{\partial m_\nu}{\partial x_\mu} = v^\mu \frac{\partial m_\mu}{\partial x_\mu} &= v^\mu \frac{\partial}{\partial x_\mu}(1-d_x^\mu)(1-d_y^\mu)(1-d_z^\mu) M_{i,\,j,\,k} + \cdots + v^\mu \frac{\partial}{\partial x_\mu} d_x^\mu d_y^\mu d_z^\mu M_{i+1,\,j+1,\,k+1} \\ &= v^\mu \left(-\frac{N_x}{L_x}\right)(1-d_y^\mu)(1-d_z^\mu) M_{i,\,j,\,k} + \cdots + v^\mu \frac{N_x}{L_x} d_y^\mu d_z^\mu M_{i+1,\,j+1,\,k+1} \quad , \end{aligned} \tag{C.4}$$

which is, similarly to the CIC paint case, a readout of the 3D field $\boldsymbol{M}$ at position $\boldsymbol{r}_\mu$, multiplied by v_μ and replacing d_x^μ with N_x/L_x and $1 - d_x^\mu$ with $-N_x/L_x$.

The VJP with respect to the 3D field $\boldsymbol{M}$ instead reads:

$$\mathrm{vjp}_{\hat{\imath},\,\hat{\jmath},\,\hat{k}} = v^\nu \frac{\partial m_\nu}{\partial M_{\hat{\imath},\,\hat{\jmath},\,\hat{k}}} \quad . \tag{C.5}$$

Since ν is summed over, all masses in positions $\boldsymbol{r}_\nu$ neighboring $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$ will bring non-zero contributions, because in a CIC readout a mesh mass $M_{i,\,j,\,k}$ contributes to all neighboring particles. Specifically, this contribution is the same that a neighboring particle at $\boldsymbol{r}_\nu$ with mass v_ν would bring to the mesh point $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$, therefore this VJP corresponds to a CIC paint of particles with masses $\boldsymbol{v}$.

C.2 VJP of Discrete Fourier Transforms (DFTs)

DFTs are simple linear transformations. Considering a real discrete input $\boldsymbol{x}$, its forward DFT can be written as:

$$\boldsymbol{y} = F\boldsymbol{x} \quad , \tag{C.6}$$

where F is the DFT matrix. The backward DFT can then be written as (ignoring normalization factors)

$$\boldsymbol{x} = F^H \boldsymbol{y} \quad , \tag{C.7}$$

where F^H is the Hermitian conjugate of F, that is $F^H = (F^*)^T = (F^T)^*$. The VJP of a forward DFT for 1D data can then be computed simply as (note that $\boldsymbol{v}$ is complex):

$$v^j \frac{dy_j}{dx_i} = v^j F_{ji} = F^T \boldsymbol{v} = \left(F^H\right)^* \boldsymbol{v} = \left(F^H \boldsymbol{v}^*\right)^* \quad , \tag{C.8}$$

that is, the complex conjugate of the backward DFT of the complex conjugate of $\boldsymbol{v}$. The backward DFT case is very similar:

$$v^j \frac{dx_j}{dy_i} = v^j F^H_{ji} = \left(F^H\right)^T \boldsymbol{v} = F^* \boldsymbol{v} = (F \boldsymbol{v}^*)^* \quad , \tag{C.9}$$

which is essentially the inverse of the above. It is easy to show that the argument is also valid for 3D data.

C.3 Calculation of the power spectrum and its VJP

Consider a field $\delta_{\boldsymbol{p}} = \delta_{u,v,w}$ in real space defined on a mesh and its DFT $\tilde{\delta}_{\boldsymbol{q}} = \tilde{\delta}_{l,m,n}$ in Fourier space. The calculation of the power spectrum of δ is performed by looping over the mesh points $\boldsymbol{q}$, computing the norm of the wavevector associated to each point $k_{\boldsymbol{q}}$, which depends on how the DFT algorithm stores its result, accumulating the contribution to the power $P(k_{\boldsymbol{q}})$ (and increasing a contributions counter $M_{k_{\boldsymbol{q}}}$), given by $|\tilde{\delta}_{\boldsymbol{q}}|^2$, and finally normalizing by the number of contributions. Additionally, one can introduce a correction to compensate for the aliasing effect brought by the usage of a MAS to paint the field onto the mesh, see for instance [283] and Appendix F. This is precisely what is done in Pylians [162], and the implementation of the calculation of the power spectrum in the `JERALD` code follows it exactly. Contrarily to Pylians, however, the power spectrum implementation in `JERALD` is trivially parallelized, where first each rank accumulates the contributions to $P(k)$ and M_k of the local part of the field only (see Appendix B), and then the local contributions $P(k)$ and mode counts M_k are exchanged among all ranks via an `MPI_Allreduce` operation, after which each power spectrum term $P(k)$ is normalized correctly.

In formula, the power spectrum of a field δ defined on a mesh of size N for a simulation box of size L (both assumed constant along each dimension for simplicity) can be written as

$$P(k) = P_k = \left(\frac{L}{N^2}\right)^3 \frac{1}{M_k} \sum_{\boldsymbol{q}} |\tilde{\delta}_{\boldsymbol{q}}|^2 \, \delta_K(\|\boldsymbol{k}_{\boldsymbol{q}}\| = k) \, \frac{1}{\tilde{W}(\boldsymbol{k}_{\boldsymbol{p}})^2} \quad , \tag{C.10}$$

where M_k indicates the number of modes for frequency k, $\delta_K(\|\boldsymbol{k}_{\boldsymbol{q}}\| = k)$ is the Kronerker delta function (since the wavevectors are discrete) and $W(\boldsymbol{k}_{\boldsymbol{q}})$ is the MAS correction factor. In what follows, $\tilde{W}(\boldsymbol{k}_{\boldsymbol{q}})^{-2}$ will be indicated as $C_{\boldsymbol{q}}$ for ease of notation.

The VJP of such a quantity is then (defining $A \equiv (L/N^2)^3$):

$$\begin{aligned}
\mathrm{vjp}_{\hat{\boldsymbol{p}}} = \sum_k v^k \frac{dP_k}{d\delta_{\hat{\boldsymbol{p}}}} &= \sum_k v^k \frac{d}{d\delta_{\hat{\boldsymbol{p}}}} A \frac{1}{M_k} \sum_{\boldsymbol{q}} C_{\boldsymbol{q}} \, |\tilde{\delta}_{\boldsymbol{q}}|^2 \, \delta_K(\|\boldsymbol{k}_{\boldsymbol{q}}\| = k) \\
&= A \sum_k v^k \frac{1}{M_k} \sum_{\boldsymbol{q}} C_{\boldsymbol{q}} \, \frac{d|\tilde{\delta}_{\boldsymbol{q}}|^2}{d\delta_{\hat{\boldsymbol{p}}}} \, \delta_K(\|\boldsymbol{k}_{\boldsymbol{q}}\| = k) \\
&= A \sum_k v^k \frac{1}{M_k} \sum_{\boldsymbol{q}} C_{\boldsymbol{q}} \, \frac{d|\tilde{\delta}_{\boldsymbol{q}}|^2}{d\tilde{\delta}_{\boldsymbol{q}}} \, \frac{d\tilde{\delta}_{\boldsymbol{q}}}{d\delta_{\hat{\boldsymbol{p}}}} \, \delta_K(\|\boldsymbol{k}_{\boldsymbol{q}}\| = k) \\
&= A \sum_{\boldsymbol{q}} C_{\boldsymbol{q}} \, \frac{v^{k_{\boldsymbol{q}}}}{M_{k_{\boldsymbol{q}}}} \, \frac{d|\tilde{\delta}_{\boldsymbol{q}}|^2}{d\tilde{\delta}_{\boldsymbol{q}}} \, \frac{d\tilde{\delta}_{\boldsymbol{q}}}{d\delta_{\hat{\boldsymbol{p}}}} \\
&= 2A \sum_{\boldsymbol{q}} C_{\boldsymbol{q}} \, \frac{v^{k_{\boldsymbol{q}}}}{M_{k_{\boldsymbol{q}}}} \, \tilde{\delta}_{\boldsymbol{q}} \, \frac{d\tilde{\delta}_{\boldsymbol{q}}}{d\delta_{\hat{\boldsymbol{p}}}} \quad .
\end{aligned} \tag{C.11}$$

Note that $\boldsymbol{v}$ and $\boldsymbol{M}$ are not defined over the entire mesh, but only for the unique wavevector moduli; accordingly, $v^{k_{\boldsymbol{q}}}$ and $M_{k_{\boldsymbol{q}}}$ indicate the values of these quantities computed for $\|\boldsymbol{k}_{\boldsymbol{q}}\|$. Using the definition of the DFT (ignoring normalizations)

$$\tilde{\delta}_{\boldsymbol{q}} = \sum_{\boldsymbol{p}} F^{\boldsymbol{p}}_{\boldsymbol{q}} \, \delta_{\boldsymbol{p}} \quad , \tag{C.12}$$

the VJP becomes

$$\begin{aligned}
\mathrm{vjp}_{\hat{\boldsymbol{p}}} &= 2A \sum_{\boldsymbol{q}} C_{\boldsymbol{q}} \, \frac{v^{k_{\boldsymbol{q}}}}{M_{k_{\boldsymbol{q}}}} \, \tilde{\delta}_{\boldsymbol{q}} \, \sum_{\boldsymbol{p}} F^{\boldsymbol{p}}_{\boldsymbol{q}} \, \frac{\partial \delta_{\boldsymbol{p}}}{\partial \delta_{\hat{\boldsymbol{p}}}} \\
&= 2A \sum_{\boldsymbol{q}} C_{\boldsymbol{q}} \, \frac{v^{k_{\boldsymbol{q}}}}{M_{k_{\boldsymbol{q}}}} \, \tilde{\delta}_{\boldsymbol{q}} \, \sum_{\boldsymbol{p}} F^{\boldsymbol{p}}_{\boldsymbol{q}} \, \delta_K(\boldsymbol{p} = \hat{\boldsymbol{p}}) \\
&= 2A \sum_{\boldsymbol{q}} C_{\boldsymbol{q}} \, \frac{v^{k_{\boldsymbol{q}}}}{M_{k_{\boldsymbol{q}}}} \, \tilde{\delta}_{\boldsymbol{q}} \, F^{\hat{\boldsymbol{p}}}_{\boldsymbol{q}} \quad ,
\end{aligned} \tag{C.13}$$

where $\delta_K(\boldsymbol{p} = \hat{\boldsymbol{p}})$ emerges because, of course, among the field values $\delta_{\boldsymbol{p}}$ in the sum of the first line, only that for which $\boldsymbol{p} = \hat{\boldsymbol{p}}$ has a dependence on $\delta_{\hat{\boldsymbol{p}}}$. Eq. (C.13) is, up to a normalization, the inverse DFT of the field $\tilde{\delta}_{\boldsymbol{q}}$ multiplied by $C_{\boldsymbol{q}}$, $v^{k_{\boldsymbol{q}}}$ and $1/M_{k_{\boldsymbol{q}}}$; assuming $P(k)$, and therefore also the M_k array, has been calculated in the forward pass, the calculation of the VJP is then straightforward. The parallelization itself is trivial, as it enters only through the FTs.

Hermitian symmetry

Note that fields δ are always real and their FTs $\tilde{\delta}$ Hermitian symmetric; therefore, for the purpose of computational efficiency, DFTs in the `JERALD` code are computed using

the FFTW method `fftw_mpi_plan_dft_r2c_3d`, which does not store $\tilde{\delta}_{l,m,n<0}$. Modes with $n < 0$ are the conjugate of those with $0 < -n < N/2$ assuming N even, therefore the inverse DFT, computed via `fftw_mpi_plan_dft_c2r_3d`, reads (ignoring normalization factors)

$$\begin{aligned}\delta_{u,v,w} = &\sum_{l,m,n=0} \mathrm{Re}\left[\tilde{\delta}_{l,m,n} F^{l,m,n}_{u,v,w}\right] + \sum_{l,m,n=N/2} \mathrm{Re}\left[\tilde{\delta}_{l,m,n} F^{l,m,n}_{u,v,w}\right] + \\ &+ \sum_{0<n<N/2} \sum_{l,m} 2\,\mathrm{Re}\left[\tilde{\delta}_{l,m,n} F^{l,m,n}_{u,v,w}\right] \quad ,\end{aligned} \tag{C.14}$$

where the last term pairs the $n < 0$ modes with the $0 < n < N/2$ ones exploiting Hermitian symmetry, so that only the field $\tilde{\delta}_{l,m,n\geq 0}$ is needed. Note that the first two terms are real by construction for a Hermitian-symmetric field; however, FFTW does not verify Hermiticity of the input field, hence the Re[·] operator. Since the calculation of the power spectrum involves an average over modes and the $n < 0$ modes are missing from $\tilde{\delta}$, it is important to either restore them or neglect *all* redundant modes $\boldsymbol{q}$ that have a conjugate $\boldsymbol{q}'$ such that $|\tilde{\delta}_{\boldsymbol{q}}|^2 = |\tilde{\delta}_{\boldsymbol{q}'}|^2$ (for example, $\boldsymbol{q} = (l < 0, 0, 0)$), so that overall each k bin in the power spectrum receives only the unique contributions of $\tilde{\delta}$, and M_k is then half of what it would be on the full field, ensuring the correct normalization. Both Pylians and the implementation of the $P(k)$ in `JERALD` use the second method. This implies that, practically, the sums in Eq. (C.13) (and (C.10) and (C.11)) are constrained to unique modes or, equivalently, $\tilde{\delta}'_{\boldsymbol{q}} = \tilde{\delta}_{\boldsymbol{q}}\, C_{\boldsymbol{q}}\, v^{k_{\boldsymbol{q}}}/M_{k_{\boldsymbol{q}}} = 0$ for redundant modes, so the field is not Hermitian symmetric and cannot be naively inverted using `fftw_mpi_plan_dft_c2r_3d`. Nonetheless, the inverse DFT of $\tilde{\delta}'_{\boldsymbol{q}}$ is

$$\begin{aligned}\delta'_{u,v,w} = &\sum_{l,m,n=0} \tilde{\delta}'_{l,m,n} F^{l,m,n}_{u,v,w} + \sum_{l,m,n=N/2} \tilde{\delta}'_{l,m,n} F^{l,m,n}_{u,v,w} + \\ &+ \sum_{0<n<N/2} \sum_{l,m} \tilde{\delta}'_{l,m,n} F^{l,m,n}_{u,v,w} + \sum_{n<0} \sum_{l,m} \tilde{\delta}'_{l,m,n} F^{l,m,n}_{u,v,w} \quad .\end{aligned} \tag{C.15}$$

Notice that, theoretically, from Eq. (C.13) $\delta'_{u,v,w}$ is real; therefore, considering the real part of Eq. (C.15) and remembering that $\tilde{\delta}'_{l,m,n<0} = 0$:

$$\begin{aligned}\mathrm{Re}\left[\delta'_{u,v,w}\right] = &\sum_{l,m,n=0} \mathrm{Re}\left[\tilde{\delta}'_{l,m,n} F^{l,m,n}_{u,v,w}\right] + \sum_{l,m,n=N/2} \mathrm{Re}\left[\tilde{\delta}'_{l,m,n} F^{l,m,n}_{u,v,w}\right] + \\ &+ \sum_{0<n<N/2} \sum_{l,m} \mathrm{Re}\left[\tilde{\delta}'_{l,m,n} F^{l,m,n}_{u,v,w}\right] \quad .\end{aligned} \tag{C.16}$$

This equation is very similar to Eq. (C.14): the inverse DFT of $\delta'_{u,v,w}$ then can still be computed using `fftw_mpi_plan_dft_c2r_3d` by simply halving $\tilde{\delta}'_{l,m,n}$ for $0 < n < N/2$.

D Numerical operators: derivatives and MAS

This appendix discusses the practical choices made in the implementation of the model in Section 2.1, focusing specifically on the evaluation of the gradient of the effective potential and the treatment of the MAS used to paint particles on the mesh and vice-versa.

D.1 Spectral and finite-difference derivatives

The displacement of Eq. (2.2) requires the calculation of the gradient of a potential known only at the mesh points $\boldsymbol{r}_g$. Two options are available: the first is to differentiate spectrally, multiplying the Fourier-space potential by $i\boldsymbol{k}$; the second is to apply a finite-difference kernel, which, being a convolution in configuration space, corresponds to a multiplication in Fourier space as well. For a centered kernel of order $2n$ acting along, say, the x axis, the operator reads

$$D_x(\boldsymbol{k}) = \frac{i}{\Delta} \sum_{m=1}^{n} c_m \sin(m k_x \Delta) \quad . \tag{D.1}$$

For the fourth- and sixth-order cases,

$$\tilde{D}^{(4)}(k_x) = \frac{i}{6\Delta}\Big[8\sin(k_x\Delta) - \sin(2k_x\Delta)\Big] = ik_x\left[1 - \frac{(k_x\Delta)^4}{30} + \mathcal{O}\big((k_x\Delta)^6\big)\right] \quad , \tag{D.2}$$

$$\begin{aligned} \tilde{D}^{(6)}(k_x) &= \frac{i}{30\Delta}\Big[45\sin(k_x\Delta) - 9\sin(2k_x\Delta) + \sin(3k_x\Delta)\Big] \\ &= ik_x\left[1 - \frac{(k_x\Delta)^6}{140} + \mathcal{O}\big((k_x\Delta)^8\big)\right] \quad . \end{aligned} \tag{D.3}$$

In the strict sense, the spectral derivative is the more principled choice: it is the exact derivative of the *band-limited* field defined of the mesh, while Eqs. (D.2) and (D.3) are approximations to it. However, this presupposes that the field on the mesh is genuinely band-limited, which it is not: it was produced by assigning a discrete set of particles to a mesh, an operation that aliases power from above the Nyquist frequency $k_N = \pi/\Delta$ back into the resolved range (see Appendices A and E). The modes closest to k_N are therefore the least reliable ones; the spectral and finite-difference operators differ most in that range: every centered difference kernel satisfies $\tilde{D}(k_N) = 0$, suppressing those modes entirely, whereas $i\boldsymbol{k}$ assigns them the largest weight of all. A finite-difference derivative thus sacrifices accuracy but gains insensitivity to the part of the spectrum most contaminated by the mass assignment. Nonetheless, the usage of both Eqs. (D.2) and (D.3) as well as ik_x itself was tested in the code, producing no appreciable difference; the likely reason is provided in Section D.3.

D.2 MAS corrections

The second choice concerns the MAS. Painting particles onto the mesh to compute the effective potential in Eq (2.5) convolves the density with the CIC kernel, so that the Fourier-space density carries a factor $\tilde{W}(\boldsymbol{k}) = \prod_i \mathrm{sinc}^2(k_i\Delta/2)$ (see Appendices E and F), and interpolating the resulting displacement back to the particle positions applies the same kernel a second time. The force experienced by the particles is therefore suppressed by $\tilde{W}^2$ relative to the one that would follow from the unsmoothed field, and a fully deconvolved scheme would divide by $\tilde{W}^2$ before the inverse transform. Nonetheless, there are two reasons to ignore this: first, $\tilde{W}$ decreases towards the Nyquist frequency, so $1/\tilde{W}^2$ amplifies precisely the aliased modes discussed above (the corrected force would then be more accurate but noisier); second, the loss of Eq. (2.10) compares two fields that have both been painted with the same kernel, so that a common factor of $\tilde{W}$ is present on either side and partially cancels in the comparison. On the contrary, in the calculation of a power spectrum, deconvolution of the MAS kernel is unambiguously appropriate and applied as standard (see Section C.3). `JERALD` was tested both with and without deconvolving, again with no noticeable performance changes, as predicted by the result in Section D.3.

D.3 Degeneracy with the Green's function

Both choices above amount to multiplying the Fourier-space displacement by a kernel that depends on k:

$$\tilde{\boldsymbol{S}}(\boldsymbol{k}) \;=\; \boldsymbol{D}(\boldsymbol{k})\,\frac{1}{\tilde{W}^{\,p}(\boldsymbol{k})}\,\tilde{G}(k)\,\tilde{f}(\boldsymbol{k}) \quad , \tag{D.4}$$

with $p = 0$ or 2 according to whether the assignment correction is applied. Considering the expression of the Green's function (see e.g. Eq. (2.7)), any factor depending on $k = \|\boldsymbol{k}\|$ alone can be reabsorbed into $\tilde{G}(k)$ during the optimization. However, the part of these factors that depends on the direction of $\boldsymbol{k}$ cannot be absorbed, since $\tilde{G}(k)$ is isotropic by construction.

For the CIC kernel this residual is very small:

$$\begin{aligned} W(\boldsymbol{k}) &\simeq \prod_i \left[1 - \frac{(k_i\Delta)^2}{24} + \frac{(k_i\Delta)^4}{1920} + \mathcal{O}\big((k_i\Delta)^6\big)\right]^2 \\ &= \prod_i \left[1 - \frac{(k_i\Delta)^2}{12} + \frac{(k_i\Delta)^4}{360} + \mathcal{O}\big((k_i\Delta)^6\big)\right] \\ &= 1 - \frac{\Delta^2}{12}\sum_i k_i^2 + \frac{\Delta^4}{360}\sum_i k_i^4 - \frac{\Delta^4}{144}\sum_{i\neq j} k_i^2 k_j^2 + \mathcal{O}\big((k_{i,\,j,\,k}\Delta)^6\big) \quad , \end{aligned} \tag{D.5}$$

where $k^6_{i,\,j,\,k}$ indicates mixed products with combined power 6; the leading term is propor-

tional to $\sum_i k_i^2 = k^2$ and is therefore isotropic: the anisotropy of the CIC kernel enters only at fourth order, through $\sum_i k_i^4$ and mixed products, which is not expressible in terms of k alone. To the accuracy at which the model operates, then, the assignment correction is essentially isotropic, and an optimized $\tilde{G}(k)$ can then absorb it practically to its full extent.

The derivative operator is more delicate, since $D(k_i)/ik_i$ depends on the individual components of $\boldsymbol{k}$ rather than on its modulus. Its magnitude, however, is small wherever the model carries appreciable weight: from Eq. (D.3), the fourth-order kernel departs from the spectral derivative by $(k_i\Delta)^4/30$, which for the configuration used in Chapter 2 (512^3 mesh, $80h^{-1}$Mpc box, $k_N \simeq 20\,h\mathrm{Mpc}^{-1}$) amounts to $\sim 3\times10^{-4}$ at $k = 2\,h\mathrm{Mpc}^{-1}$ and to about 1% at $k = 5\,h\mathrm{Mpc}^{-1}$, while in the case of Chapter 3 (128^3 mesh, $25h^{-1}$Mpc box, $k_N \simeq 16\,h\mathrm{Mpc}^{-1}$) is $\sim 8\times10^{-4}$ at $k = 2\,h\mathrm{Mpc}^{-1}$ and $\sim 3\%$ at $k = 5\,h\mathrm{Mpc}^{-1}$. Since the fitted Green's functions suppress the displacement at large wavenumbers through both the exponential cutoff and the power-law factor (Eq. (2.7)), the region in which the two operators differ appreciably contributes little to the displacement in the first place. These estimates are consistent with what is found empirically: using either Eq. (D.2), Eq. (D.3) or the spectral derivative, and applying or omitting the deconvolution of the assignment kernel, produce no significant change in the performance of the model.

Interestingly, this analysis implies that the fitted $\tilde{G}(k)$ absorbs the numerical transfer functions of the scheme alongside the sub-grid physics it is meant to describe, and is therefore not a purely physical object. Moreover, quantities built from $\tilde{G}(k)$ (in particular the combination $J(k)$ of Eq. (3.3)) inherit a dependence on Δ, and therefore on mesh size and box size; comparisons of $J(k)$ across configurations with different resolutions should then account for this.

E Role of $J(k)$ and γ in the loss

The purpose of this section is to show how, under specific assumptions, $J(k)$ and γ enter in the loss that `JERALD` optimizes. The calculation itself may be considered unfinished at the time of writing of this thesis, as developing it further could give additional insights; nonetheless, what follows still provides useful results for the discussion of Chapter 3. Consider the case of a simulation with box size L and a mesh of size N. In what follows, ρ will always indicate a normalized density, i.e. a density divided by its mean, or $\delta + 1$, where δ is the overdensity. `JERALD`'s loss is defined as an L_2 norm of the difference between the normalized CAMELS and `JERALD` densities computed through CIC on a mesh:

$$\mathcal{L} = \|\rho^{\text{cic}}_{\text{camels}}(\boldsymbol{r}_g) - \rho^{\text{cic}}_{\texttt{JERALD}}(\boldsymbol{r}_g)\|^2 \quad ,$$

where $\boldsymbol{r}_g$ are discretized positions $\boldsymbol{r}_g = \Delta\boldsymbol{n}$, where $\Delta = L/N$ is the mesh spacing and $\boldsymbol{n}$ is a 3D integer vector, with $n_i = 0, \ldots, N-1$. Now, using Parseval's theorem:

$$\mathcal{L} = \|\tilde{\rho}^{\text{cic,dft}}_{\text{camels}}(\boldsymbol{k}) - \tilde{\rho}^{\text{cic,dft}}_{\texttt{JERALD}}(\boldsymbol{k})\|^2 \quad ,$$

where $\boldsymbol{k}$ are discretized wavevectors $\boldsymbol{k} = 2\pi/L\,\boldsymbol{m}$, where once again $\boldsymbol{m}$ is a 3D integer vector with $m_i = 0, \ldots, N-1$ and the superscript "dft" indicates the fact that these are Discrete Fourier Transforms (DFTs).

A generic density in real (continuous) space defined for N_p particles with masses M_p in a periodic box reads

$$\rho'(\boldsymbol{r}) = \sum_{p=0}^{N_p-1} M_p\,\delta(\boldsymbol{r} - \boldsymbol{r}_p) \quad ,$$

where $\boldsymbol{r}_p$ are the particles positions and $\delta(\boldsymbol{r})$ is the Dirac delta over the finite volume of the simulation box, defined as

$$\delta(\boldsymbol{r}) = \frac{1}{L^3}\sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \quad ,$$

where the wavevectors $\boldsymbol{k} = 2\pi/L\,\boldsymbol{m}$, $\boldsymbol{m} \in \mathbb{Z}^3$. The normalized density is then:

$$\rho(\boldsymbol{r}) = \frac{\rho'(\boldsymbol{r})}{\bar{\rho}'} = \frac{L^3}{M_{tot}}\rho'(\boldsymbol{r}) = \frac{L^3}{M_{tot}}\sum_p M_p\,\delta(\boldsymbol{r} - \boldsymbol{r}_p) \quad , \qquad M_{tot} = \sum_p M_p \quad .$$

Now define $A := L^3/M_{tot}$ for ease of notation. CIC interpolation practically corresponds

to a convolution between the density field and the CIC kernel

$$W(\boldsymbol{r}) = \frac{1}{\Delta^3} \prod_{i=x,y,z} \max\left(0, 1 - \frac{|r_i|}{\Delta}\right) \quad ,$$

i.e.

$$\rho^{\rm cic}(\boldsymbol{r}) = \int_V d\boldsymbol{r}'\, W(\boldsymbol{r} - \boldsymbol{r}')\, \rho(\boldsymbol{r}') = A \sum_p M_p\, W(\boldsymbol{r} - \boldsymbol{r}_p) \quad .$$

The FT of this quantity is then:

$$\begin{aligned}
\tilde{\rho}^{\rm cic}(\boldsymbol{k}) &= \int_V d\boldsymbol{r} \rho^{\rm cic}(\boldsymbol{r})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \\
&= A \int_V d\boldsymbol{r} \sum_p M_p\, W(\boldsymbol{r} - \boldsymbol{r}_p)\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \\
&= A \sum_p M_p \int_V d\boldsymbol{r}\, W(\boldsymbol{r} - \boldsymbol{r}_p)\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \\
&= A\, \tilde{W}(\boldsymbol{k}) \sum_p M_p\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_p}
\end{aligned}$$

Technically, this sum could be evaluated explicitly for an arbitrary number of wavevectors $N_{\boldsymbol{k}}$, resulting in an alias-free Fourier-space density, although it is very computationally expensive, scaling as $\mathcal{O}(N_p \times N_{\boldsymbol{k}})$. In practice, as it is customary doing, `JERALD` uses FFT algorithms to compute an (aliased) estimate of $\tilde{\rho}^{\rm cic}(\boldsymbol{r})$; this requires interpolating $\rho^{\rm cic}(\boldsymbol{r})$ onto a mesh, which corresponds to evaluating it at the mesh points $\boldsymbol{r}_g$ mentioned above:

$$\rho^{\rm cic}(\boldsymbol{r}_g) = A \sum_p M_p\, W(\boldsymbol{r}_g - \boldsymbol{r}_p) \quad ;$$

the DFT of this quantity is then (using forward normalization to match the `JERALD` code)

$$\begin{aligned}
\tilde{\rho}^{\rm cic,dft}(\boldsymbol{k}) &= \frac{1}{N^3} \sum_g \rho^{\rm cic}(\boldsymbol{r}_g) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_g} \\
&= \frac{A}{N^3} \sum_g \sum_p M_p\, W(\boldsymbol{r}_g - \boldsymbol{r}_p) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_g} \quad .
\end{aligned}$$

This quantity is better understood if one introduces the so-called sampling function [85]

$$\text{III}(\boldsymbol{r}) = \sum_{\boldsymbol{n}} \delta(\boldsymbol{r} - \boldsymbol{n}\Delta) \quad ,$$

where $\boldsymbol{n}$ is an integer vector with $n_i = 0, \ldots, N-1$ (The same function is often written with $\boldsymbol{n} \in \mathbb{Z}^3$, which in unbounded space is the standard definition; in a periodic box the

two coincide), so that

$$\begin{aligned}\tilde{\rho}^{\mathrm{cic,dft}}(\boldsymbol{k}) &= \frac{A}{N^3}\sum_p M_p \sum_g W(\boldsymbol{r}_g - \boldsymbol{r}_p) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_g} \\ &= \frac{A}{N^3}\sum_p M_p \int_V d\boldsymbol{r}\, \mathrm{III}(\boldsymbol{r}) W(\boldsymbol{r} - \boldsymbol{r}_p) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \quad .\end{aligned}$$

The integral represents the FT of the product between the sampling function and the CIC kernel, which corresponds to a convolution (see appendices F and G for the explicit calculations of the FTs of $W(\boldsymbol{r})$ and $\mathrm{III}(\boldsymbol{r})$):

$$\begin{aligned}\tilde{\rho}^{\mathrm{cic,dft}}(\boldsymbol{k}) &= \frac{A}{N^3}\sum_p M_p \int_V d\boldsymbol{r}\, \mathrm{III}(\boldsymbol{r}) W(\boldsymbol{r} - \boldsymbol{r}_p) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \\ &= A\sum_p M_p \sum_{\boldsymbol{k}'} \left(\sum_{\boldsymbol{n}} \delta_{\boldsymbol{k}',\, 2k_N\boldsymbol{n}} \right) \tilde{W}(\boldsymbol{k} - \boldsymbol{k}') e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}_p} \\ &= A\sum_{\boldsymbol{n}} \tilde{W}(\boldsymbol{k} - 2k_N\boldsymbol{n}) \sum_p M_p\, e^{-i(\boldsymbol{k}-2k_N\boldsymbol{n})\cdot\boldsymbol{r}_p} \\ &= \sum_{\boldsymbol{n}} \tilde{\rho}^{\mathrm{cic}}(\boldsymbol{k} - 2k_N\boldsymbol{n}) \quad ,\end{aligned}$$

where $k_N = \pi/\Delta$ is the Nyquist frequency of the mesh. This is a standard result, in which the term $\boldsymbol{n} = \boldsymbol{0}$ corresponds to the true FT of the CIC density field (including the effect of the interpolation), while the other contributions are aliasing artifacts of the usage of a DFT. The size of these terms is controlled by the CIC kernel itself. As Appendix F shows,

$$\tilde{W}(\boldsymbol{k}) = \prod_{i=x,y,z} \left(\frac{\sin k_i\Delta/2}{k_i\Delta/2} \right)^2 \quad ,$$

and shifting the argument by a reciprocal-lattice vector $\boldsymbol{k} \to \boldsymbol{k} - 2k_N\boldsymbol{n}$ changes the sine only by a sign, $\sin(k_i\Delta/2 - \pi n_i) = (-1)^{n_i}\sin(k_i\Delta/2)$. Therefore the contributions of the aliasing terms compared to the one at $\boldsymbol{n} = \boldsymbol{0}$ is reduced by a factor:

$$\frac{\tilde{W}(\boldsymbol{k} - 2k_N\boldsymbol{n})}{\tilde{W}(\boldsymbol{k})} = \prod_i \left(\frac{k_i\Delta/2}{k_i\Delta/2 - \pi n_i} \right)^2 \xrightarrow[\|\boldsymbol{k}\| \ll k_N]{} \prod_i \left(\frac{k_i}{2k_N n_i} \right)^2 \quad ,$$

the limit being taken over the non-vanishing components of $\boldsymbol{n}$. The leading alias is therefore suppressed as $(k/2k_N)^2$, where $k = \|\boldsymbol{k}\|$: with the $L = 25h^{-1}\mathrm{Mpc}$, $N = 128$ grids used in Chapter 3, $2k_N \simeq 32\, h\mathrm{Mpc}^{-1}$, so the correction is $\sim 4\times10^{-3}$ at $k \simeq 2\, h\mathrm{Mpc}^{-1}$ and $\sim 2\times10^{-2}$ at $k \simeq 5\, h\mathrm{Mpc}^{-1}$, becoming appreciable only as k approaches the Nyquist frequency itself. Over the range of scales on which the argument below is used these

contributions are thus negligible, and will accordingly be ignored, so that $\tilde{\rho}^{\mathrm{cic,dft}} = \tilde{\rho}^{\mathrm{cic}}$.

Focusing on the density computed by `JERALD`, the particles positions can be written as $\boldsymbol{r}_p = \boldsymbol{r}_p^{\mathrm{fpm}} + \boldsymbol{S}_p$, where $\boldsymbol{r}_p^{\mathrm{fpm}}$ are the FastPM positions and $\boldsymbol{S}_p$ the `JERALD` displacements. Since the $\boldsymbol{S}_p$ are small, the exponential in the Fourier-space density can be expanded:

$$\begin{aligned}
\tilde{\rho}^{\mathrm{cic}}_{\texttt{JERALD}}(\boldsymbol{k}) &= A\tilde{W}(\boldsymbol{k}) \sum_p M_p \, e^{-i\boldsymbol{k}\cdot(\boldsymbol{r}_p^{\mathrm{fpm}} + \boldsymbol{S}_p)} \\
&\approx A\tilde{W}(\boldsymbol{k}) \sum_p M_p \, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_p^{\mathrm{fpm}}} \, (1 - i\boldsymbol{k}\cdot\boldsymbol{S}_p) \\
&= \tilde{\rho}^{\mathrm{cic}}_{\mathrm{fpm}}(\boldsymbol{k}) - iA\tilde{W}(\boldsymbol{k}) \, \boldsymbol{k} \cdot \sum_p M_p \boldsymbol{S}_p \, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_p^{\mathrm{fpm}}} \quad .
\end{aligned}$$

The displacement is read out at the input particle positions with the same CIC kernel:

$$\boldsymbol{S}_p = \Delta^3 \sum_g W(\boldsymbol{r}_g - \boldsymbol{r}_p^{\mathrm{fpm}}) \boldsymbol{S}(\boldsymbol{r}_g) \quad ,$$

$\boldsymbol{S}(\boldsymbol{r}_g)$ being the displacement field evaluated on the mesh. Substituting into the previous expression and swapping the two sums,

$$\begin{aligned}
\boldsymbol{I}(\boldsymbol{k}) = A \sum_p M_p \, \boldsymbol{S}_p \, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_p^{\mathrm{fpm}}} &= A \, \Delta^3 \sum_g \boldsymbol{S}(\boldsymbol{r}_g) \sum_p M_p \, W(\boldsymbol{r}_g - \boldsymbol{r}_p^{\mathrm{fpm}}) \, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_p^{\mathrm{fpm}}} \\
= A \, \Delta^3 \sum_g e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_g} \, \boldsymbol{S}(\boldsymbol{r}_g) \sum_p M_p \, W(\boldsymbol{r}_g - \boldsymbol{r}_p^{\mathrm{fpm}}) \, e^{-i\boldsymbol{k}\cdot(\boldsymbol{r}_p^{\mathrm{fpm}} - \boldsymbol{r}_g)} \quad .
\end{aligned}$$

Expanding the phase factor at the end of this expression and retaining only the zeroth-order term neglects a correction that is second order in $k\Delta$, since $W(\boldsymbol{u})$ is even and the density varies smoothly on the scale of a cell; for the grids used here it amounts to $\sim 1\%$ at $k \simeq 2\,h\mathrm{Mpc}^{-1}$. It will thus be ignored in what follows, consistently with the large-scale regime in which the argument is applied, and is in any case a smooth, deterministic factor rather than a source of scatter, so that the conclusions in Chapter 3 do not change. The term $\boldsymbol{I}(\boldsymbol{k})$ then becomes:

$$\begin{aligned}
\boldsymbol{I}(\boldsymbol{k}) &\simeq A \, \Delta^3 \sum_g e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_g} \, \boldsymbol{S}(\boldsymbol{r}_g) \sum_p M_p \, W(\boldsymbol{r}_g - \boldsymbol{r}_p^{\mathrm{fpm}}) \\
&= \Delta^3 \sum_g e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_g} \, \boldsymbol{S}(\boldsymbol{r}_g) \, \rho^{\mathrm{cic}}_{\mathrm{fpm}}(\boldsymbol{r}_g) = L^3 \, \mathrm{DFT}\big[\boldsymbol{S} \, \rho^{\mathrm{cic}}_{\mathrm{fpm}}\big](\boldsymbol{k}) \quad ,
\end{aligned}$$

where the last equality uses the DFT normalization convention adopted above. The displacement field itself is constructed entirely on the mesh: following Eq. (2.2), the gridded field $\rho^{\mathrm{cic}}_{\mathrm{fpm}}$ is raised to the power γ to form the source $f = \big(\rho^{\mathrm{cic}}_{\mathrm{fpm}}\big)^{\gamma}$, transformed, multiplied by the Green's function and by the gradient kernel, and transformed back.

Since in the implementation the gradient is evaluated through a finite-difference kernel rather than spectrally, this gives

$$\tilde{\boldsymbol{S}}(\boldsymbol{k}) = i\boldsymbol{D}(\boldsymbol{k})\,\tilde{G}(k)\,\tilde{f}(\boldsymbol{k}) \simeq i\boldsymbol{k}\,\tilde{G}(k)\,\tilde{f}(\boldsymbol{k}) \quad ,$$

where the finite-difference kernel $\boldsymbol{D}(\boldsymbol{k})$ can be approximated to $\boldsymbol{k}$ for the scales considered here, as its Fourier representation is $k_i\Big[1+\mathcal{O}\big((k_i\Delta)^4\big)\Big]$, and it departs from $i\boldsymbol{k}$ by less than 10^{-3} at $k \simeq 2\,h\mathrm{Mpc}^{-1}$ and by about three per cent at $k \simeq 5\,h\mathrm{Mpc}^{-1}$ (see Section D.1). The transform of the product of the two fields defined on the mesh is exactly a discrete convolution of their transforms, giving

$$\Delta\tilde{\rho}(\boldsymbol{k}) \equiv \tilde{\rho}^{\mathrm{cic}}_{\texttt{JERALD}}(\boldsymbol{k}) - \tilde{\rho}^{\mathrm{cic}}_{\mathrm{fpm}}(\boldsymbol{k}) = L^3\tilde{W}(\boldsymbol{k})\sum_{\boldsymbol{q}}(\boldsymbol{k}\cdot\boldsymbol{q})\,\tilde{G}(q)\,\tilde{f}(\boldsymbol{q})\,\tilde{\rho}^{\mathrm{cic}}_{\mathrm{fpm}}(\boldsymbol{k}-\boldsymbol{q}) \quad ,$$

where the two factors of i have canceled. Already at this stage the structure of the degeneracy is manifest: the model enters only through the product $\tilde{G}(q)\tilde{f}(\boldsymbol{q})$, never through $\tilde{G}(q)$ or its parameters alone. Since $\tilde{G}(q)$ is a function of the modulus $\|\boldsymbol{q}\|$ only, it is isotropic and carries no information about phases; the source transform $\tilde{f}(\boldsymbol{q})$, on the other hand, carries both an amplitude and a phase. Separating the two by introducing

$$u(\boldsymbol{q}) \equiv \frac{\tilde{f}(\boldsymbol{q})}{\sqrt{P_{\rho^\gamma}(q)}} \quad , \qquad \big\langle|u|^2\big\rangle_{\|\boldsymbol{q}\|=q} = 1 \ \text{ by construction,}$$

gives

$$\tilde{G}(q)\,\tilde{f}(\boldsymbol{q}) = J(q)\,u(\boldsymbol{q}) \quad , \qquad J(q) \equiv \tilde{G}(q)\sqrt{P_{\rho^\gamma}(q)} \quad ,$$

with $J(q)$ as in Eq. (3.3). This decomposition is not one choice among many: because $\tilde{G}(q)$ depends on the modulus alone, $J(q)$ is the unique combination for which the remaining factor has unit power in every shell, any other isotropic rescaling leaving part of that amplitude behind. The model can therefore affect the prediction in exactly two ways: through $J(q)$, which sets the amplitude of the correction as a function of scale, and through $u(\boldsymbol{q})$, which retains everything else about the source field (i.e. its phases, and the fluctuations of its mode amplitudes about the shell average) and hence determines how the resulting displacement is arranged in space relative to the structures of the input. The parameters α, k_l and ν enter the former alone; γ, on the other hand, operates on both. Defining $R(\boldsymbol{k}) = \tilde{\rho}^{\mathrm{cic}}_{\mathrm{camels}}(\boldsymbol{k}) - \tilde{\rho}^{\mathrm{cic}}_{\mathrm{fpm}}(\boldsymbol{k})$ as the residual that the displacement is required to

supply, the loss reads

$$\begin{aligned}\mathcal{L} &= \sum_{\boldsymbol{k}} |R(\boldsymbol{k}) - \Delta\tilde{\rho}(\boldsymbol{k})|^2 \\ &= \sum_{\boldsymbol{k}} |R(\boldsymbol{k})|^2 - 2\,\mathrm{Re}\sum_{\boldsymbol{k}} R^*(\boldsymbol{k})\Delta\tilde{\rho}(\boldsymbol{k}) + \sum_{\boldsymbol{k}} |\Delta\tilde{\rho}(\boldsymbol{k})|^2 \quad ,\end{aligned}$$

with

$$\Delta\tilde{\rho}(\boldsymbol{k}) = L^3\tilde{W}(\boldsymbol{k}) \sum_{\boldsymbol{q}} (\boldsymbol{k}\cdot\boldsymbol{q})\, J(q)\, u(\boldsymbol{q})\, \tilde{\rho}^{\mathrm{cic}}_{\mathrm{fpm}}(\boldsymbol{k}-\boldsymbol{q}) \quad .$$

The first term is a constant, independent of the model. The last is quadratic, and expanding it produces a double sum over $\boldsymbol{q}$ and $\boldsymbol{q}'$ of the form $\sum_{\boldsymbol{q},\boldsymbol{q}'} J(q)J(q')\, u(\boldsymbol{q})u^*(\boldsymbol{q}')\, C(\boldsymbol{q},\boldsymbol{q}')$, where C is built entirely from the geometry and from the input field. It is thus a quadratic form in $J(q)$, and introduces no dependence on the model beyond $J(q)$ and the phases of u already identified above. Reducing it to a closed expression, however, is not trivial: the diagonal terms $\boldsymbol{q} = \boldsymbol{q}'$ are treated straightforwardly, but the off-diagonal ones $\boldsymbol{q} \neq \boldsymbol{q}'$ are not; this is one of the aspects in which the calculation presented here remains incomplete, as anticipated at the beginning of this appendix, and it is left for future work.

The cross term is worth examining in more detail, because it is linear in the model and it makes the residual dependence on γ explicit. Defining

$$b(\boldsymbol{k},\, \boldsymbol{q}) \equiv \frac{\tilde{f}(\boldsymbol{q})\, \tilde{\rho}^{\mathrm{cic}}_{\mathrm{fpm}}(\boldsymbol{k}-\boldsymbol{q})\, R^*(\boldsymbol{k})}{\sqrt{P_{\rho^\gamma}(q)}} \quad , \tag{E.1}$$

which involves only the input and target fields together with γ, that term becomes

$$-2\sum_{\boldsymbol{k},\boldsymbol{q}} \tilde{W}(\boldsymbol{k})\, (\boldsymbol{k}\cdot\boldsymbol{q})\, J(q)\, \mathrm{Re}\, b(\boldsymbol{k},\boldsymbol{q}) \quad .$$

Note that b is a three-point object, correlating two modes of the input field with one mode of the residual, and is therefore sensitive to the spatial alignment between the source term and the discrepancy the displacement is required to remove. Raising a field to a power is a non-linear pointwise operation, and thus changing γ alters both the amplitude and the phases of the source modes; the division by $\sqrt{P_{\rho^\gamma}(q)}$ removes amplitude dependence, which is absorbed into $J(q)$; however, the phase information remains, allowing γ to be recovered reproducibly across optimizations on the same target, which is observed in Figure 3.17. The same fact implies that any summary of the input field which discards phase information, the power spectrum included, is structurally unable to account for that part of γ. This is the prediction discussed in Section 3.5, and the reason a regressor operating on the full input field is suggested there as the natural next step.

Finally, it is worth remembering that the expansion of the loss in this appendix is incomplete, as it lacks the treatment of the quadratic term, and with it the explicit determination of the $J(k)$ at which the loss is stationary: evaluating that expression for a given FastPM/CAMELS pair and comparing it with the $J(k)$ recovered by the optimizations would be a direct test of the calculations presented here.

F Calculation of the FT of the CIC kernel

This appendix shows the explicit calculation of the Fourier Transform of the Cloud-In-Cell kernel.

$$
\begin{aligned}
\tilde{W}(\boldsymbol{k}) &= \int d\boldsymbol{r}\, W(\boldsymbol{r}) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \\
&= \prod_{i=x,y,z} \int dr_i\, W(r_i) e^{-ik_i r_i} \\
&= \frac{1}{\Delta^3} \prod_{i=x,y,z} \int_{-\Delta}^{\Delta} dr_i \left(1 - \frac{|r_i|}{\Delta}\right) e^{-ik_i r_i} \\
&= \frac{1}{\Delta^3} \prod_{i=x,y,z} \left[\int_{-\Delta}^{0} dr_i \left(1 + \frac{r_i}{\Delta}\right) e^{-ik_i r_i} + \int_{0}^{\Delta} dr_i \left(1 - \frac{r_i}{\Delta}\right) e^{-ik_i r_i}\right] \\
&= \frac{1}{\Delta^3} \prod_{i=x,y,z} \left[\int_{0}^{\Delta} dr_i \left(1 - \frac{r_i}{\Delta}\right) e^{ik_i r_i} + \int_{0}^{\Delta} dr_i \left(1 - \frac{r_i}{\Delta}\right) e^{-ik_i r_i}\right] \\
&= \frac{1}{\Delta^3} \prod_{i=x,y,z} \int_{0}^{\Delta} dr_i \left(1 - \frac{r_i}{\Delta}\right) \left(e^{ik_i r_i} + e^{-ik_i r_i}\right) \\
&= \frac{2}{\Delta^3} \prod_{i=x,y,z} \int_{0}^{\Delta} dr_i \left(1 - \frac{r_i}{\Delta}\right) \cos k_i r_i \\
&= \frac{2}{\Delta^3} \prod_{i=x,y,z} \left[\left.\frac{\sin k_i r_i}{k_i}\right|_0^{\Delta} - \left.\frac{\cos k_i r_i + k_i r_i \sin k_i r_i}{k_i^2 \Delta}\right|_0^{\Delta}\right] \\
&= \frac{4}{\Delta^3} \prod_{i=x,y,z} \left[-\frac{1}{k_i^2 \Delta} \frac{\cos k_i \Delta - 1}{2}\right] = \prod_{i=x,y,z} \left(\frac{\sin k_i \Delta/2}{k_i \Delta/2}\right)^2 \quad .
\end{aligned}
$$

G Calculation of the FT of the sampling function

This appendix shows the explicit calculation of the Fourier Transform of the sampling function [85].

$$\begin{aligned}
\tilde{\text{III}}(\boldsymbol{k}) &= \int_V d\boldsymbol{r} \sum_{\boldsymbol{n}} \delta(\boldsymbol{r} - \boldsymbol{n}\Delta) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \\
&= \sum_{\boldsymbol{n}} e^{-i\boldsymbol{k}\cdot\boldsymbol{n}\Delta} \\
&= \prod_{i=x,y,z} \sum_{n_i} e^{-ik_i n_i \Delta} \\
&= \prod_{i=x,y,z} \frac{1 - e^{-ik_i N\Delta}}{1 - e^{-ik_i\Delta}} = \prod_{i=x,y,z} \frac{1 - e^{-ik_i L}}{1 - e^{-ik_i L/N}} \\
&= \prod_{i=x,y,z} \frac{e^{-ik_i L/2}}{e^{-ik_i L/2N}} \frac{\sin k_i L/2}{\sin k_i L/2N} \quad .
\end{aligned}$$

Now, since the simulation box is periodic, the wavevectors are constrained to $\boldsymbol{k} = 2\pi/L\,\boldsymbol{m}$, where $\boldsymbol{m}$ is an integer vector:

$$\begin{aligned}
\tilde{\text{III}}(\boldsymbol{k}) &= \prod_{i=x,y,z} \frac{e^{-ik_i L/2}}{e^{-ik_i L/2N}} \frac{\sin k_i L/2}{\sin k_i L/2N} \\
&= \prod_{i=x,y,z} \frac{e^{-i\pi\, m_i}}{e^{-i\pi\, m_i/N}} \frac{\sin \pi\, m_i}{\sin \pi\, m_i/N} \\
&= N^3 \sum_{\boldsymbol{n}} \delta_{\boldsymbol{m},\, N\boldsymbol{n}} = N^3 \sum_{\boldsymbol{n}} \delta_{\boldsymbol{k},\, 2k_N\boldsymbol{n}} \quad ,
\end{aligned}$$

where $k_N = \pi/\Delta$ is the Nyquist frequency for the mesh.

Bibliography

[1] Mauro Rigo, Roberto Trotta, and Matteo Viel. Jerald: high-fidelity dark matter, stellar mass, and neutral hydrogen maps from fast n-body simulations. *Monthly Notices of the Royal Astronomical Society*, 541(1):166–178, June 2025.

[2] Giulio Scelfo, Satvik Mishra, Mauro Rigo, Roberto Trotta, and Matteo Viel. Field-level multi-tracers simulation-based inference of cosmological parameters from 3d maps. *arXiv preprint arXiv:2605.26210*, 2026.

[3] P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

[4] Scott Dodelson and Fabian Schmidt. *Modern Cosmology*. 2020.

[5] N Aghanim et al. Planck 2018 results. vi. cosmological parameters. *Astron. Astrophys*, 641:A6, 2020.

[6] George F Smoot, Charles L Bennett, Alan Kogut, Edward L Wright, John Aymon, Nancy W Boggess, Edward S Cheng, Giovanni De Amici, Steven Gulkis, Michael G Hauser, et al. Structure in the cobe differential microwave radiometer first-year maps. *Astrophysical Journal, Part 2-Letters (ISSN 0004-637X), vol. 396, no. 1, Sept. 1, 1992, p. L1-L5. Research supported by NASA.*, 396:L1–L5, 1992.

[7] Planck Collaboration. Planck 2018 results. I. Overview and the cosmological legacy of Planck. *A&A*, 641:A1, September 2020.

[8] David W Hogg, Daniel J Eisenstein, Michael R Blanton, Neta A Bahcall, Jon Brinkmann, James E Gunn, and Donald P Schneider. Cosmic homogeneity demonstrated with luminous red galaxies. *The Astrophysical Journal*, 624(1):54–58, 2005.

[9] Morag I Scrimgeour, Tamara Davis, Chris Blake, J Berian James, Gregory B Poole, Lister Staveley-Smith, Sarah Brough, Matthew Colless, Carlos Contreras, Warrick Couch, et al. The wigglez dark energy survey: the transition to large-scale cosmic homogeneity. *Monthly Notices of the Royal Astronomical Society*, 425(1):116–134, 2012.

[10] Albert Einstein. Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pages 142–152, January 1917.

[11] Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the national academy of sciences*, 15(3):168–173, 1929.

[12] A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386, January 1922.

[13] A. Friedmann. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift fur Physik*, 21(1):326–332, December 1924.

[14] Georges Lemaître. Un univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles, A47, p. 49-59*, 47:49–59, 1927.

[15] Fritz Zwicky. On the masses of nebulae and of clusters of nebulae. *The Astrophysical Journal*, 86:217, 1937.

[16] Vera C Rubin and W Kent Ford Jr. Rotation of the andromeda nebula from a spectroscopic survey of emission regions. *Astrophysical Journal, vol. 159, p. 379*, 159:379, 1970.

[17] Vera C Rubin, W Kent Ford Jr, and Norbert Thonnard. Rotational properties of 21 sc galaxies with a large range of luminosities and radii, from ngc 4605/r= 4kpc/to ugc 2885/r= 122 kpc. *Astrophysical Journal, Part 1, vol. 238, June 1, 1980, p. 471-487.*, 238:471–487, 1980.

[18] Douglas Clowe, Maruša Bradač, Anthony H Gonzalez, Maxim Markevitch, Scott W Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal Letters*, 648(2):L109–L113, 2006.

[19] Timothy MC Abbott, Michel Aguena, Alex Alarcon, S Allam, O Alves, A Amon, F Andrade-Oliveira, James Annis, S Avila, D Bacon, et al. Dark energy survey year 3 results: Cosmological constraints from galaxy clustering and weak lensing. *Physical Review D*, 105(2):023520, 2022.

[20] Daniel J Eisenstein, Idit Zehavi, David W Hogg, Roman Scoccimarro, Michael R Blanton, Robert C Nichol, Ryan Scranton, Hee-Jong Seo, Max Tegmark, Zheng Zheng, et al. Detection of the baryon acoustic peak in the large-scale correlation

function of sdss luminous red galaxies. *The Astrophysical Journal*, 633(2):560–574, 2005.

[21] Shaun Cole, Will J Percival, John A Peacock, Peder Norberg, Carlton M Baugh, Carlos S Frenk, Ivan Baldry, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, et al. The 2df galaxy redshift survey: power-spectrum analysis of the final data set and cosmological implications. *Monthly Notices of the Royal Astronomical Society*, 362(2):505–534, 2005.

[22] Gianfranco Bertone and Dan Hooper. History of dark matter. *Reviews of Modern Physics*, 90(4):045002, 2018.

[23] Richard H Cyburt, Brian D Fields, Keith A Olive, and Tsung-Han Yeh. Big bang nucleosynthesis: Present status. *Reviews of Modern Physics*, 88(1):015004, 2016.

[24] Dale J Fixsen. The temperature of the cosmic microwave background. *The Astrophysical Journal*, 707(2):916–920, 2009.

[25] Adam G Riess, Alexei V Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M Garnavich, Ron L Gilliland, Craig J Hogan, Saurabh Jha, Robert P Kirshner, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The astronomical journal*, 116(3):1009–1038, 1998.

[26] Saul Perlmutter, Goldhaber Aldering, Gerson Goldhaber, Richard A Knop, Peter Nugent, Patricia G Castro, Susana Deustua, Sebastien Fabbro, Ariel Goobar, Donald E Groom, et al. Measurements of ω and λ from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565–586, 1999.

[27] M Abdul Karim, J Aguilar, Steven Ahlen, Shadab Alam, Lori Allen, C Allende Prieto, Otávio Alves, Abhijeet Anand, Uendert Andrade, E Armengaud, et al. Desi dr2 results. ii. measurements of baryon acoustic oscillations and cosmological constraints. *Physical Review D*, 112(8):083515, 2025.

[28] Kushal Lodha, R Calderon, WL Matthewson, Arman Shafieloo, Mustapha Ishak, Jiaming Pan, C Garcia-Quintero, D Huterer, Georgios Valogiannis, LA Ureña-López, et al. Extended dark energy analysis using desi dr2 bao measurements. *Physical Review D*, 112(8):083511, 2025.

[29] Steven Weinberg. The cosmological constant problem. *Reviews of modern physics*, 61(1):1, 1989.

[30] P James E Peebles and Bharat Ratra. The cosmological constant and dark energy. *Reviews of modern physics*, 75(2):559, 2003.

[31] Edmund J Copeland, Mohammad Sami, and Shinji Tsujikawa. Dynamics of dark energy. *arXiv preprint hep-th/0603057*, 2006.

[32] Nick Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 1987.

[33] Andrew JS Hamilton. Linear redshift distortions: A review. *The Evolving Universe: Selected Topics on Large-Scale Structure and on the Properties of Galaxies*, pages 185–275, 1998.

[34] Ya B Zel'Dovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and Astrophysics, Vol. 5, p. 84-89*, 5:84–89, 1970.

[35] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, Jan 1981.

[36] Viatcheslav F Mukhanov and Gennady V Chibisov. Quantum fluctuations and a nonsingular universe. *ZhETF Pisma Redaktsiiu*, 33:549–553, 1981.

[37] Juan Maldacena. Non-gaussian features of primordial fluctuations in single field inflationary models. *Journal of High Energy Physics*, 2003(05):013–013, 2003.

[38] Yashar Akrami, Frederico Arroja, M Ashdown, J Aumont, Carlo Baccigalupi, M Ballardini, Anthony J Banday, RB Barreiro, Nicola Bartolo, S Basak, et al. Planck 2018 results. x. constraints on inflation. *arXiv preprint arXiv:1807.06211*, 2018.

[39] James M Bardeen, JR Bond, Nick Kaiser, and Albert S Szalay. The statistics of peaks of gaussian random fields. *The Astrophysical Journal*, 304:15, 1986.

[40] Daniel J Eisenstein and Wayne Hu. Baryonic features in the matter transfer function. *The Astrophysical Journal*, 496(2):605–614, 1998.

[41] D. Blas, J. Lesgourgues, and T. Tram. The cosmic linear anisotropy solving system (class). *JCAP*, 2011.

[42] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models. *The Astrophysical Journal*, 538(2):473–476, 2000.

[43] Philip JE Peebles and JT Yu. Primeval adiabatic perturbation in an expanding universe. *Astrophysical Journal, vol. 162, p. 815*, 162:815, 1970.

[44] Rashid A Sunyaev and Ya B Zeldovich. Small-scale fluctuations of relic radiation. *Astrophysics and Space Science*, 7(1):3–19, 1970.

[45] A Meiksin, Martin White, and JA Peacock. Baryonic signatures in large-scale structure. *Monthly Notices of the Royal Astronomical Society*, 304(4):851–864, 1999.

[46] John Huchra, Marc Davis, David Latham, and John Tonry. A survey of galaxy redshifts. iv-the data. *Astrophysical Journal Supplement Series (ISSN 0067-0049), vol. 52, June 1983, p. 89-119. Research supported by the Smithsonian Institution.*, 52:89–119, 1983.

[47] Valérie De Lapparent, Margaret J Geller, and John P Huchra. A slice of the universe. *Astrophysical Journal, Part 2-Letters to the Editor (ISSN 0004-637X), vol. 302, March 1, 1986, p. L1-L5. Research supported by the Smithsonian Institution.*, 302:L1–L5, 1986.

[48] Avishai Dekel and Ofer Lahav. Stochastic nonlinear galaxy biasing. *The Astrophysical Journal*, 520(1):24–34, 1999.

[49] Vincent Desjacques, Donghui Jeong, and Fabian Schmidt. Large-scale galaxy bias. *Physics reports*, 733:1–193, 2018.

[50] Hiroo Totsuji and Taro Kihara. The correlation function for the distribution of galaxies. *Publications of the Astronomical Society of Japan*, 21(3):221–229, 1969.

[51] P. J. E. Peebles. The Gravitational-Instability Picture and the Nature of the Distribution of Galaxies. *ApJL*, 189:L51, April 1974.

[52] Edward J Groth and PJE Peebles. Statistical analysis of catalogs of extragalactic objects. vii-two-and three-point correlation functions for the high-resolution shane-wirtanen catalog of galaxies. *Astrophysical Journal, Part 1, vol. 217, Oct. 15, 1977, p. 385-405. Research supported by the Alfred P. Sloan Foundation and NSF.*, 217:385–405, 1977.

[53] Chris Blake and Karl Glazebrook. Probing dark energy using baryonic oscillations in the galaxy power spectrum as a cosmological ruler. *The Astrophysical Journal*, 594(2):665–673, 2003.

[54] Hee-Jong Seo and Daniel J Eisenstein. Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys. *The Astrophysical Journal*, 598(2):720–740, 2003.

[55] Donald G York, Jennifer Adelman, John E Anderson, Jr, Scott F Anderson, James Annis, Neta A Bahcall, JA Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, et al. The sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3):1579–1587, 2000.

[56] Kyle S Dawson, David J Schlegel, Christopher P Ahn, Scott F Anderson, Éric Aubourg, Stephen Bailey, Robert H Barkhouser, Julian E Bautista, Alessandra Beifiori, Andreas A Berlind, et al. The baryon oscillation spectroscopic survey of sdss-iii. *The Astronomical Journal*, 145(1):10, 2013.

[57] Kyle S Dawson, Jean-Paul Kneib, Will J Percival, Shadab Alam, Franco D Albareti, Scott F Anderson, Eric Armengaud, Éric Aubourg, Stephen Bailey, Julian E Bautista, et al. The sdss-iv extended baryon oscillation spectroscopic survey: overview and early data. *The Astronomical Journal*, 151(2):44, 2016.

[58] DESI Collaboration. The DESI Experiment Part I: Science, Targeting, and Survey Design. *arXiv e-prints*, page arXiv:1611.00036, October 2016.

[59] Jelte TA de Jong, Gijs A Verdoes Kleijn, Konrad H Kuijken, Edwin A Valentijn, KiDS, and Astro-WISE Consortiums. The kilo-degree survey. *Experimental Astronomy*, 35(1):25–44, 2013.

[60] Dark Energy Survey Collaboration:, T Abbott, FB Abdalla, J Aleksić, S Allam, A Amara, D Bacon, E Balbinot, M Banerji, K Bechtol, et al. The dark energy survey: more than dark energy–an overview. *Monthly Notices of the Royal Astronomical Society*, 460(2):1270–1299, 2016.

[61] Hiroaki Aihara, Nobuo Arimoto, Robert Armstrong, Stéphane Arnouts, Neta A Bahcall, Steven Bickerton, James Bosch, Kevin Bundy, Peter L Capak, James HH Chan, et al. The hyper suprime-cam ssp survey: overview and survey design. *Publications of the Astronomical Society of Japan*, 70(SP1):S4, 2018.

[62] Luca et al. Amendola. Cosmology and fundamental physics with the euclid satellite. *Living Reviews in Relativity*, 21(1), April 2018.

[63] Željko Ivezić et al. Lsst: From science drivers to reference design and anticipated data products. *The Astrophysical Journal*, 873(2):111, March 2019.

[64] Matthias Bartelmann and Peter Schneider. Weak gravitational lensing. *Physics Reports*, 340(4-5):291–472, 2001.

[65] Martin Kilbinger. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901, 2015.

[66] SKA Cosmology SWG. Cosmology with Phase 1 of the Square Kilometre Array: Red Book 2018: Technical specifications and performance forecasts. *Publ. Astron. Soc. Austral.*, 37:e007, 2020.

[67] Steven R Furlanetto, S Peng Oh, and Frank H Briggs. Cosmology at low frequencies: The 21 cm transition and the high-redshift universe. *Physics reports*, 433(4-6):181–301, 2006.

[68] Jonathan R Pritchard and Abraham Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, 2012.

[69] Philip Bull, Pedro G Ferreira, Prina Patel, and Mario G Santos. Late-time cosmology with 21 cm intensity mapping experiments. *The Astrophysical Journal*, 803(1):21, 2015.

[70] Rupert AC Croft, David H Weinberg, Neal Katz, and Lars Hernquist. Recovery of the power spectrum of mass fluctuations from observations of the lyα forest. *The Astrophysical Journal*, 495(1):44–62, 1998.

[71] Michael Rauch. The lyman alpha forest in the spectra of quasistellar objects. *Annual Review of Astronomy and Astrophysics*, 36(1):267–316, 1998.

[72] Matthew McQuinn. The evolution of the intergalactic medium. *Annual Review of Astronomy and Astrophysics*, 54(1):313–362, 2016.

[73] Robert E Smith, JA Peacock, A Jenkins, SDM White, CS Frenk, FR Pearce, Peter A Thomas, G Efstathiou, and HMP Couchman. Stable clustering, the halo model and non-linear cosmological power spectra. *Monthly Notices of the Royal Astronomical Society*, 341(4):1311–1332, 2003.

[74] Ryuichi Takahashi, Masanori Sato, Takahiro Nishimichi, Atsushi Taruya, and Masamune Oguri. Revising the halofit model for the nonlinear matter power spectrum. *The Astrophysical Journal*, 761(2):152, 2012.

[75] Annalisa Pillepich, Volker Springel, Dylan Nelson, Shy Genel, Jill Naiman, Rüdiger Pakmor, Lars Hernquist, Paul Torrey, Mark Vogelsberger, Rainer Weinberger, and Federico Marinacci. Simulating galaxy formation with the illustristng model. *Monthly Notices of the Royal Astronomical Society*, 473(3):4077–4106, 01 2018.

[76] Volker Springel. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Monthly Notices of the Royal Astronomical Society*, 401(2):791–851, 01 2010.

[77] James S. Bolton, Ewald Puchwein, Debora Sijacki, Martin G. Haehnelt, Tae-Sun Kim, Avery Meiksin, John A. Regan, and Matteo Viel. The sherwood simulation suite: overview and data comparisons with the lyman-α forest at redshifts $2 \leq z \leq 5$. *Monthly Notices of the Royal Astronomical Society*, 464(1):897–914, 01 2017.

[78] Ewald Puchwein, James S Bolton, Laura C Keating, Margherita Molaro, Prakash Gaikwad, Girish Kulkarni, Martin G Haehnelt, Vid Iršič, Tomáš Šoltinský, Matteo Viel, Dominique Aubert, George D Becker, and Avery Meiksin. The sherwood–relics simulations: overview and impact of patchy reionization and pressure smoothing on the intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 519(4):6162–6183, 03 2023.

[79] Volker Springel. The cosmological simulation code gadget-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134, 12 2005.

[80] Dylan Nelson, Annalisa Pillepich, Volker Springel, Rainer Weinberger, Lars Hernquist, Rüdiger Pakmor, Shy Genel, Paul Torrey, Mark Vogelsberger, Guinevere Kauffmann, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: the galaxy colour bimodality. *Monthly Notices of the Royal Astronomical Society*, 475(1):624–647, 11 2017.

[81] Rachel S Somerville and Romeel Davé. Physical models of galaxy formation in a cosmological framework. *Annual Review of Astronomy and Astrophysics*, 53(1):51–113, 2015.

[82] Thorsten Naab and Jeremiah P Ostriker. Theoretical challenges in galaxy formation. *Annual review of astronomy and astrophysics*, 55:59–109, 2017.

[83] Mark Vogelsberger, Federico Marinacci, Paul Torrey, and Ewald Puchwein. Cosmological simulations of galaxy formation. *Nature Reviews Physics*, 2(1):42–66, 2020.

[84] Josh Barnes and Piet Hut. A hierarchical o (n log n) force-calculation algorithm. *nature*, 324(6096):446–449, 1986.

[85] Roger W Hockney and James W Eastwood. *Computer simulation using particles*. crc Press, 2021.

[86] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[87] Svetlin Tassev, Matias Zaldarriaga, and Daniel J. Eisenstein. Solving large scale structure in ten easy steps with cola. *Journal of Cosmology and Astroparticle Physics*, 2013(06):036, jun 2013.

[88] Albert Izard, Martin Crocce, and Pablo Fosalba. Ice-cola: towards fast and accurate synthetic galaxy catalogues optimizing a quasi-n-body method. *Monthly Notices of the Royal Astronomical Society*, 459(3):2327–2341, 2016.

[89] Yu Feng, Man-Yat Chu, Uros Seljak, and Patrick McDonald. Fastpm: a new scheme for fast simulations of dark matter and haloes. *Monthly Notices of the Royal Astronomical Society*, 463(3):2273–2286, 08 2016.

[90] Biwei Dai, Yu Feng, and Uroš Seljak. A gradient based method for modeling baryons and matter in halos of fast simulations. *Journal of Cosmology and Astroparticle Physics*, 2018(11):009–009, 2018.

[91] Denise Lanzieri, François Lanusse, and Jean-Luc Starck. Hybrid physical-neural odes for fast n-body simulations. *arXiv preprint arXiv:2207.05509*, 2022.

[92] Siyu He, Yin Li, Yu Feng, Shirley Ho, Siamak Ravanbakhsh, Wei Chen, and Barnabás Póczos. Learning to predict the cosmological structure formation. *Proceedings of the National Academy of Sciences*, 116(28):13825–13832, 2019.

[93] Drew Jamieson, Yin Li, Renan Alves de Oliveira, Francisco Villaescusa-Navarro, Shirley Ho, and David N. Spergel. Field-level Neural Network Emulator for Cosmological N-body Simulations. *The Astrophysical Journal*, 952(2):145, July 2023. Publisher: The American Astronomical Society.

[94] Francis Bernardeau, Stéphane Colombi, Enrique Gaztañaga, and Román Scoccimarro. Large-scale structure of the universe and cosmological perturbation theory. *Physics reports*, 367(1-3):1–248, 2002.

[95] Takahiko Matsubara. Resumming cosmological perturbations via the lagrangian picture: One-loop results in real space and in redshift space. *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, 77(6):063530, 2008.

[96] FR Bouchet, S Colombi, E Hivon, and R Juszkiewicz. Perturbative lagrangian approach to gravitational instability. *arXiv preprint astro-ph/9406013*, 1994.

[97] Thomas Buchert. Lagrangian theory of gravitational instability of friedman–lemaître cosmologies–a generic third-order model for non-linear clustering. *Monthly Notices of the Royal Astronomical Society*, 267(4):811–820, 1994.

[98] Paolo Catelan. Lagrangian dynamics in non-flat universes and non-linear gravitational evolution. *Monthly Notices of the Royal Astronomical Society*, 276(1):115–124, 1995.

[99] Martín Crocce, Sebastián Pueblas, and Román Scoccimarro. Transients from initial conditions in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 373(1):369–381, 2006.

[100] Patrick McDonald and Arabindo Roy. Clustering of dark matter tracers: generalizing bias for the coming era of precision lss. *Journal of Cosmology and Astroparticle Physics*, 2009(08):020–020, 2009.

[101] Valentin Assassi, Daniel Baumann, Daniel Green, and Matias Zaldarriaga. Renormalized halo bias. *Journal of Cosmology and Astroparticle Physics*, 2014(08):056–056, 2014.

[102] Takahiko Matsubara. Resumming cosmological perturbations via the lagrangian picture: One-loop results in real space and in redshift space. *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, 77(6):063530, 2008.

[103] Francisco Villaescusa-Navarro, Shy Genel, Emanuele Castorina, Andrej Obuljen, David N Spergel, Lars Hernquist, Dylan Nelson, Isabella P Carucci, Annalisa Pillepich, Federico Marinacci, et al. Ingredients for 21 cm intensity mapping. *The Astrophysical Journal*, 866(2):135, 2018.

[104] Andreu Arinyo-i Prats, Jordi Miralda-Escudé, Matteo Viel, and Renyue Cen. The non-linear power spectrum of the lyman alpha forest. *Journal of Cosmology and Astroparticle Physics*, 2015(12):017–017, 2015.

[105] Chirag Modi, Shi-Fan Chen, and Martin White. Simulations and symmetries. *Monthly Notices of the Royal Astronomical Society*, 492(4):5754–5763, 2020.

[106] Nickolas Kokron, Joseph DeRose, Shi-Fan Chen, Martin White, and Risa H Wechsler. The cosmology dependence of galaxy clustering and lensing from a hybrid n-body–perturbation theory model. *Monthly Notices of the Royal Astronomical Society*, 505(1):1422–1440, 2021.

[107] Matteo Zennaro, Raul E Angulo, Marcos Pellejero-Ibáñez, Jens Stücker, Sergio Contreras, and Giovanni Aricò. The bacco simulation project: biased tracers in real space. *Monthly Notices of the Royal Astronomical Society*, 524(2):2407–2419, 2023.

[108] Andreas A Berlind and David H Weinberg. The halo occupation distribution: Toward an empirical determination of the relation between galaxies and mass. *The Astrophysical Journal*, 575(2):587–616, 2002.

[109] Zheng Zheng, Andreas A Berlind, David H Weinberg, Andrew J Benson, Carlton M Baugh, Shaun Cole, Romeel Davé, Carlos S Frenk, Neal Katz, and Cedric G Lacey. Theoretical models of the halo occupation distribution: Separating central and satellite galaxies. *The Astrophysical Journal*, 633(2):791–809, 2005.

[110] Xiaohu Yang, HJ Mo, and Frank C van den Bosch. Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies. *Monthly Notices of the Royal Astronomical Society*, 339(4):1057–1080, 2003.

[111] Antonio Vale and JP Ostriker. Linking halo mass to galaxy luminosity. *Monthly Notices of the Royal Astronomical Society*, 353(1):189–200, 2004.

[112] Charlie Conroy, Risa H Wechsler, and Andrey V Kravtsov. Modeling luminosity-dependent galaxy clustering through cosmic time. *The Astrophysical Journal*, 647(1):201–214, 2006.

[113] Peter S Behroozi, Charlie Conroy, and Risa H Wechsler. A comprehensive analysis of uncertainties affecting the stellar mass–halo mass relation for $0 < z < 4$. *The Astrophysical Journal*, 717(1):379–403, 2010.

[114] Benjamin P. Moster, Rachel S. Somerville, Christian Maulbetsch, Frank C. van den Bosch, Andrea V. Macciò, Thorsten Naab, and Ludwig Oser. Constraints on the relationship between stellar mass and halo mass at low and high redshift. *The Astrophysical Journal*, 710(2):903, jan 2010.

[115] Benjamin P. Moster, Thorsten Naab, and Simon D. M. White. Galactic star formation and accretion histories from matching galaxies to dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 428(4):3121–3138, 12 2012.

[116] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292:371–394, May 1985.

[117] Rachel S Somerville and Joel R Primack. Semi-analytic modelling of galaxy formation: the local universe. *Monthly Notices of the Royal Astronomical Society*, 310(4):1087–1110, 1999.

[118] Andreas A. Berlind and David H. Weinberg. The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. *The Astrophysical Journal*, 575(2):587–616, August 2002.

[119] Volker Springel, Simon D. M. White, Adrian Jenkins, Carlos S. Frenk, Naoki Yoshida, Liang Gao, Julio Navarro, Robert Thacker, Darren Croton, John Helly, John A. Peacock, Shaun Cole, Peter Thomas, Hugh Couchman, August Evrard, Jörg Colberg, and Frazer Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435(7042):629–636, Jun 2005.

[120] Darren J Croton, Volker Springel, Simon DM White, Gabriella De Lucia, Carlos S Frenk, Liang Gao, Adrian Jenkins, Guinevere Kauffmann, JF Navarro, and Naoki Yoshida. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Monthly Notices of the Royal Astronomical Society*, 365(1):11–28, 2006.

[121] C. M. Baugh. A primer on hierarchical galaxy formation: the semi-analytical approach. *Reports on Progress in Physics*, 69(12):3101–3156, December 2006.

[122] Darren J. Croton, Adam R. H. Stevens, Chiara Tonini, Thibault Garel, Maksym Bernyk, Antonio Bibiano, Luke Hodkinson, Simon J. Mutch, Gregory B. Poole, and Genevieve M. Shattow. Semi-analytic galaxy evolution (sage): Model calibration and basic results. *The Astrophysical Journal Supplement Series*, 222(2):22, feb 2016.

[123] Peter Behroozi, Risa H Wechsler, Andrew P Hearin, and Charlie Conroy. UniverseMachine: The correlation between galaxy growth and dark matter halo assembly from $z = 0 - 10$. *Monthly Notices of the Royal Astronomical Society*, 488(3):3143–3194, 05 2019.

[124] Andrey V Kravtsov, Alexey Vikhlinin, and Daisuke Nagai. A new robust low-scatter x-ray mass indicator for clusters of galaxies. *The Astrophysical Journal*, 650(1):128–136, 2006.

[125] Aurel Schneider and Romain Teyssier. A new method to quantify the effects of baryons on the matter power spectrum. *Journal of Cosmology and Astroparticle Physics*, 2015(12):049–049, 2015.

[126] Giovanni Aricò, Raul E Angulo, Carlos Hernández-Monteagudo, Sergio Contreras, Matteo Zennaro, Marcos Pellejero-Ibañez, and Yetli Rosas-Guevara. Modelling the large-scale mass density field of the universe as a function of cosmology and baryonic physics. *Monthly Notices of the Royal Astronomical Society*, 495(4):4800–4819, 2020.

[127] M Arnaud, GW Pratt, R Piffaretti, H Böhringer, JH Croston, and E Pointecouteau. The universal galaxy cluster pressure profile from a representative sample of nearby systems (rexcess) and the y sz–m 500 relation. *Astronomy & Astrophysics*, 517:A92, 2010.

[128] Nicholas Battaglia, J Richard Bond, Christoph Pfrommer, and JL Sievers. On the cluster physics of sunyaev–zel'dovich and x-ray surveys. ii. deconstructing the thermal sz power spectrum. *The Astrophysical Journal*, 758(2):75, 2012.

[129] Michael Rauch, Jordi Miralda-Escude, Wallace LW Sargent, Tom A Barlow, David H Weinberg, Lars Hernquist, Neal Katz, Renyue Cen, and Jeremiah P Ostriker. The opacity of the lyα forest and implications for Ω_b and the ionizing background. *The Astrophysical Journal*, 489(1):7–20, 1997.

[130] Pierluigi Monaco. Approximate methods for the generation of dark matter halo catalogs in the age of precision cosmology. *Galaxies*, 4(4):53, 2016.

[131] Xiaoying Xu, Shirley Ho, Hy Trac, Jeff Schneider, Barnabas Poczos, and Michelle Ntampaka. A first look at creating mock catalogs with machine learning techniques. *The Astrophysical Journal*, 772(2):147, jul 2013.

[132] Harshil M. Kamdar, Matthew J. Turk, and Robert J. Brunner. Machine learning and cosmological simulations – I. Semi-analytical models. *Monthly Notices of the Royal Astronomical Society*, 455(1):642–658, 11 2015.

[133] Harshil M. Kamdar, Matthew J. Turk, and Robert J. Brunner. Machine learning and cosmological simulations – II. Hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 457(2):1162–1179, 02 2016.

[134] Shankar Agarwal, Romeel Davé, and Bruce A Bassett. Painting galaxies into dark matter haloes using machine learning. *Monthly Notices of the Royal Astronomical Society*, 478(3):3410–3422, 05 2018.

[135] Divij Sharma, Biwei Dai, Francisco Villaescusa-Navarro, and Uroš Seljak. A field-level emulator for modelling baryonic effects across hydrodynamic simulations. *MNRAS*, 538(3):1415–1426, April 2025.

[136] Xinyue Zhang, Yanfang Wang, Wei Zhang, Yueqiu Sun, Siyu He, Gabriella Contardo, Francisco Villaescusa-Navarro, and Shirley Ho. From dark matter to galaxies with convolutional networks. *arXiv preprint arXiv:1902.05965*, 2019.

[137] Peter Harrington, Mustafa Mustafa, Max Dornfest, Benjamin Horowitz, and Zarija Lukić. Fast, high-fidelity lyα forests with convolutional neural networks. *The Astrophysical Journal*, 929(2):160, apr 2022.

[138] Urmila Chadayammuri, Michelle Ntampaka, John ZuHone, Ákos Bogdán, and Ralph P Kraft. Painting baryons on to N-body simulations of galaxy clusters with image-to-image deep learning. *Monthly Notices of the Royal Astronomical Society*, 526(2):2812–2829, 09 2023.

[139] Nathanaël Perraudin, Ankit Srivastava, Aurelien Lucchi, Tomasz Kacprzak, Thomas Hofmann, and Alexandre Réfrégier. Cosmological n-body simulations: a

challenge for scalable generative models. *Computational Astrophysics and Cosmology*, 6(1):5, 2019.

[140] Tilman Tröster, Cameron Ferguson, Joachim Harnois-Déraps, and Ian G McCarthy. Painting with baryons: augmenting N-body simulations with gas using deep generative models. *Monthly Notices of the Royal Astronomical Society: Letters*, 487(1):L24–L29, 05 2019.

[141] M Bernardini, R Feldmann, D Anglés-Alcázar, M Boylan-Kolchin, J Bullock, L Mayer, and J Stadel. From EMBER to FIRE: predicting high resolution baryon fields from dark matter simulations with deep learning. *Monthly Notices of the Royal Astronomical Society*, 509(1):1323–1341, 10 2021.

[142] Benjamin Horowitz, Max Dornfest, Zarija Lukić, and Peter Harrington. hyphy: Deep generative conditional posterior mapping of hydrodynamical physics. *The Astrophysical Journal*, 941(1):42, dec 2022.

[143] Tanner Sether. Toward a deep learning approach for fast galaxy generation. In *International Conference on Machine Learning for Astrophysics*, 2024.

[144] Satvik Mishra, Roberto Trotta, and Matteo Viel. Large, fast, and accurate hi intensity maps with latent overlap diffusion. *Monthly Notices of the Royal Astronomical Society*, 545(3):staf2071, 11 2025.

[145] Satvik Mishra, Roberto Trotta, and Matteo Viel. Cosmo-fold: Fast generation and upscaling of field-level cosmological maps with overlap latent diffusion. *arXiv preprint arXiv:2601.14377*, 2026.

[146] Benjamin Horowitz, Carolina Cuesta-Lazaro, and Omar Yehia. Baryonbridge: Stochastic interpolant model for fast hydrodynamical simulations. *arXiv preprint arXiv:2510.19224*, 2025.

[147] Sultan Hassan, Francisco Villaescusa-Navarro, Benjamin Wandelt, David N. Spergel, Daniel Anglés-Alcázar, Shy Genel, Miles Cranmer, Greg L. Bryan, Romeel Davé, Rachel S. Somerville, Michael Eickenberg, Desika Narayanan, Shirley Ho, and Sambatra Andrianomena. HIFlow: Generating Diverse HI Maps and Inferring Cosmology while Marginalizing over Astrophysics using Normalizing Flows. *The Astrophysical Journal*, 937(2):83, October 2022. arXiv:2110.02983 [astro-ph].

[148] Sambatra Andrianomena, Sultan Hassan, and Francisco Villaescusa-Navarro. Cosmological multifield emulator. *Astrophysics and Space Science*, 370(5):44, 2025.

[149] Biwei Dai and Uroš Seljak. Learning effective physical laws for generating cosmological hydrodynamics with lagrangian deep learning. *Proceedings of the National Academy of Sciences*, 118(16):e2020324118, 2021.

[150] Marcel P. van Daalen, Ian G. McCarthy, and Joop Schaye. Exploring the effects of galaxy formation on matter clustering through a library of simulation power spectra. *Monthly Notices of the Royal Astronomical Society*, 491(2):2424–2446, January 2020.

[151] Marcel P. van Daalen, Joop Schaye, C. M. Booth, and Claudio Dalla Vecchia. The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology. *Monthly Notices of the Royal Astronomical Society*, 415(4):3649–3665, August 2011.

[152] Aurel Schneider, Romain Teyssier, Joachim Stadel, Nora Elisa Chisari, Amandine M. C. Le Brun, Adam Amara, and Alexandre Refregier. Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation. *Journal of Cosmology and Astroparticle Physics*, 2019(3):020, March 2019.

[153] Gabriele Parimbelli, Matteo Viel, and Emiliano Sefusatti. On the degeneracy between baryon feedback and massive neutrinos as probed by matter clustering and weak lensing. *Journal of Cosmology and Astroparticle Physics*, 2019(1):010, January 2019.

[154] Francisco Villaescusa-Navarro, Shy Genel, Emanuele Castorina, Andrej Obuljen, David N. Spergel, Lars Hernquist, Dylan Nelson, Isabella P. Carucci, Annalisa Pillepich, Federico Marinacci, Benedikt Diemer, Mark Vogelsberger, Rainer Weinberger, and Rüdiger Pakmor. Ingredients for 21 cm Intensity Mapping. *The Astrophysical Journal*, 866(2):135, October 2018.

[155] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018.

[156] Matteo Frigo and Steven G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".

[157] Wenzel Jakob, Jason Rhinelander, and Dean Moldovan. pybind11 – seamless operability between c++11 and python, 2017. https://github.com/pybind/pybind11.

[158] Dion Häfner and Filippo Vicentini. mpi4jax: Zero-copy mpi communication of jax arrays. *Journal of Open Source Software*, 6(65):3419, 2021.

[159] Vid Iršič, Matteo Viel, Martin G. Haehnelt, James S. Bolton, Margherita Molaro, Ewald Puchwein, Elisa Boera, George D. Becker, Prakash Gaikwad, Laura C. Keating, and Girish Kulkarni. Unveiling dark matter free streaming at the smallest scales with the high redshift Lyman-alpha forest. *Phys. Rev. D*, 109(4):043511, February 2024.

[160] Volker Springel. The cosmological simulation code gadget-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134, 12 2005.

[161] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[162] Francisco Villaescusa-Navarro. Pylians: Python libraries for the analysis of numerical simulations. Astrophysics Source Code Library, record ascl:1811.008, November 2018.

[163] Monica Barrera, Volker Springel, Simon D. M. White, César Hernández-Aguayo, Lars Hernquist, Carlos Frenk, Rüdiger Pakmor, Fulvio Ferlito, Boryana Hadzhiyska, Ana Maria Delgado, Rahul Kannan, and Sownak Bose. The MillenniumTNG Project: semi-analytic galaxy formation models on the past lightcone. *MNRAS*, 525(4):6312–6335, November 2023.

[164] Isabella P. Carucci, Francisco Villaescusa-Navarro, Matteo Viel, and Andrea Lapi. Warm dark matter signatures on the 21cm power spectrum: intensity mapping forecasts for SKA. *Journal of Cosmology and Astroparticle Physics*, 2015(7):047–047, July 2015.

[165] Paul Bode, Jeremiah P. Ostriker, and Neil Turok. Halo formation in warm dark matter models. *The Astrophysical Journal*, 556(1):93, jul 2001.

[166] David H. Weinberg, James S. Bullock, Fabio Governato, Rachel Kuzio de Naray, and Annika H. G. Peter. Cold dark matter: Controversies on small scales. *Proceedings of the National Academy of Science*, 112(40):12249–12255, October 2015.

[167] Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488(3):4440–4458, 07 2019.

[168] T Lucas Makinen, Tom Charnock, Pablo Lemos, Natalie Porqueres, Alan F Heavens, and Benjamin D Wandelt. The cosmic graph: Optimal information extraction from

large-scale structure using catalogues. *The Open Journal of Astrophysics*, 5(1), December 2022.

[169] T. L. Makinen, Alan Heavens, Natalia Porqueres, Tom Charnock, Axel Lapel, and Benjamin D. Wandelt. Hybrid summary statistics: neural weak lensing inference beyond the power spectrum. *Journal of Cosmology and Astroparticle Physics*, 2025(01):095, jan 2025.

[170] Tom Charnock, Guilhem Lavaux, and Benjamin D. Wandelt. Automatic physical inference with information maximising neural networks. *Physical Review D*, 97, 2018.

[171] Uroš Seljak, Grigor Aslanyan, Yu Feng, and Chirag Modi. Towards optimal extraction of cosmological information from nonlinear data. *Journal of Cosmology and Astroparticle Physics*, 2017(12), 12 2017.

[172] Biwei Dai and Uroš Seljak. Translation and rotation equivariant normalizing flow (trenf) for optimal cosmological analysis. *Monthly Notices of the Royal Astronomical Society*, 516(2):2363–2373, 07 2022.

[173] Natalia Porqueres, Alan Heavens, Daniel Mortlock, and Guilhem Lavaux. Bayesian forward modelling of cosmic shear data. *Monthly Notices of the Royal Astronomical Society*, 502(2):3035–3044, 01 2021.

[174] Carl De Boor and Carl De Boor. *A practical guide to splines*, volume 27. springer New York, 1978.

[175] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[176] Tijmen Tieleman. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26, 2012.

[177] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 2006.

[178] Marcel P van Daalen, Joop Schaye, CM Booth, and Claudio Dalla Vecchia. The effects of galaxy formation on the matter power spectrum: A challenge for precision cosmology. *Monthly Notices of the Royal Astronomical Society*, 415(4):3649–3665, 2011.

[179] Marcel P van Daalen, Ian G McCarthy, and Joop Schaye. Exploring the effects of galaxy formation on matter clustering through a library of simulation power spectra. *Monthly Notices of the Royal Astronomical Society*, 491(2):2424–2446, 2020.

[180] Robert Reischke, Dennis Neumann, Klara Antonia Bertmann, Steffen Hagstotz, and Hendrik Hildebrandt. Calibrating baryonic feedback with weak lensing and fast radio bursts, 2024.

[181] N E Chisari, M L A Richardson, J Devriendt, Y Dubois, A Schneider, A M C Le Brun, R S Beckmann, S Peirani, A Slyz, and C Pichon. The impact of baryons on the matter power spectrum from the horizon-agn cosmological hydrodynamical simulation. *Monthly Notices of the Royal Astronomical Society*, 480(3):3962–3977, 08 2018.

[182] Francisco Villaescusa-Navarro, Daniel Anglés-Alcázar, Shy Genel, David N. Spergel, Rachel S. Somerville, Romeel Dave, Annalisa Pillepich, Lars Hernquist, Dylan Nelson, Paul Torrey, Desika Narayanan, Yin Li, Oliver Philcox, Valentina La Torre, Ana Maria Delgado, Shirley Ho, Sultan Hassan, Blakesley Burkhart, Digvijay Wadekar, Nicholas Battaglia, Gabriella Contardo, and Greg L. Bryan. The camels project: Cosmology and astrophysics with machine-learning simulations. *The Astrophysical Journal*, 915(1):71, jul 2021.

[183] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, Luke Kelley, Mark Lovell, Benedikt Diemer, and Lars Hernquist. The illustristng simulations: Public data release, 2021.

[184] Annalisa Pillepich, Dylan Nelson, Lars Hernquist, Volker Springel, Rüdiger Pakmor, Paul Torrey, Rainer Weinberger, Shy Genel, Jill P Naiman, Federico Marinacci, and Mark Vogelsberger. First results from the illustristng simulations: the stellar mass content of groups and clusters of galaxies. *Monthly Notices of the Royal Astronomical Society*, 475(1):648–675, December 2017.

[185] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.

[186] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[187] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[188] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[189] DeepMind, Igor Babuschkin, Kate Baumli, Alison Bell, Surya Bhupatiraju, Jake Bruce, Peter Buchlovsky, David Budden, Trevor Cai, Aidan Clark, Ivo Danihelka, Antoine Dedieu, Claudio Fantacci, Jonathan Godwin, Chris Jones, Ross Hemsley, Tom Hennigan, Matteo Hessel, Shaobo Hou, Steven Kapturowski, Thomas Keck, Iurii Kemaev, Michael King, Markus Kunesch, Lena Martens, Hamza Merzic, Vladimir Mikulik, Tamara Norman, George Papamakarios, John Quan, Roman Ring, Francisco Ruiz, Alvaro Sanchez, Laurent Sartran, Rosalia Schneider, Eren Sezener, Stephen Spencer, Srivatsan Srinivasan, Miloš Stanojević, Wojciech Stokowiec, Luyu Wang, Guangyao Zhou, and Fabio Viola. The DeepMind JAX Ecosystem, 2020.

[190] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[191] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[192] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[193] Roberto Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.*, 49:71–104, 2008.

[194] Bartlett, Deaglan J., Wandelt, Benjamin D., Zennaro, Matteo, Ferreira, Pedro G., and Desmond, Harry. Syren-halofit: A fast, interpretable, high-precision formula for the nonlinear matter power spectrum. *A&A*, 686:A150, 2024.

[195] Florent Leclercq and Alan Heavens. On the accuracy and precision of correlation functions and field-level inference in cosmology. *Monthly Notices of the Royal Astronomical Society: Letters*, 506(1):L85–L90, 09 2021.

[196] Natalia Porqueres, Alan Heavens, Daniel Mortlock, and Guilhem Lavaux. Lifting weak lensing degeneracies with a field-based likelihood. *Monthly Notices of the Royal Astronomical Society*, 509(3):3194–3202, 01 2022.

[197] ChangHoon Hahn, Florian Beutler, Manodeep Sinha, Andreas Berlind, Shirley Ho, and David W Hogg. Likelihood non-gaussianity in large-scale structure analyses. *Monthly Notices of the Royal Astronomical Society*, 485(2):2956–2969, 05 2019.

[198] Robin E Upham, Michael L Brown, and Lee Whittaker. Sufficiency of a gaussian power spectrum likelihood for accurate cosmology from upcoming weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 503(2):1999–2013, 05 2021.

[199] Samuel Brieden, Héctor Gil-Marín, and Licia Verde. Shapefit: extracting the power spectrum shape information in galaxy surveys beyond bao and rsd. *Journal of Cosmology and Astroparticle Physics*, 2021(12):054, dec 2021.

[200] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

[201] Jan-Matthis Lueckmann, Jan Boelts, David Greenberg, Pedro Goncalves, and Jakob Macke. Benchmarking simulation-based inference. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 343–351. PMLR, 13–15 Apr 2021.

[202] Raul E. Angulo and Oliver Hahn. Large-scale dark matter simulations. *Living Reviews in Computational Astrophysics*, 8(1):1, December 2022. arXiv:2112.05165 [astro-ph].

[203] Niall Jeffrey, Justin Alsing, and François Lanusse. Likelihood-free inference with neural compression of DES SV weak lensing map statistics. *Monthly Notices of the Royal Astronomical Society*, 501(1):954–969, December 2020. arXiv:2009.08459 [astro-ph].

[204] Konstantin Karchev, Roberto Trotta, and Christoph Weniger. Simsims: Simulation-based supernova ia model selection with thousands of latent variables, 2023.

[205] Florian List, Noemi Anau Montel, and Christoph Weniger. Bayesian simulation-based inference for cosmological initial conditions. *arXiv preprint arXiv:2310.19910*, 2023.

[206] Anchal Saxena, P Daniel Meerburg, Christoph Weniger, Eloy de Lera Acedo, and Will Handley. Simulation-based inference of the sky-averaged 21-cm signal from cd-eor with reach. *RAS Techniques and Instruments*, 3(1):724–736, January 2024.

[207] Rahul Srinivasan, Enrico Barausse, Natalia Korsakova, and Roberto Trotta. Simulation-based population inference of LISA's Galactic binaries: Bypassing the global fit. *Phys. Rev. D*, 112(10):103043, 2025.

[208] Anna Lueber, Konstantin Karchev, Chloe Fisher, Matthias Heim, Roberto Trotta, and Kevin Heng. Near-instantaneous atmospheric retrievals and model comparison with faster. *The Astrophysical Journal Letters*, 984(1):L32, apr 2025.

[209] Francesco Sinigaglia, Patricia Iglesias-Navarro, and Matteo Viel. Simulation-based inference from the lyman-alpha forest 1d power spectrum with camels. *arXiv preprint arXiv:2603.13011*, 2026.

[210] Pauline Gorbatchev, Jean-Luc Starck, Stefano Camera, and Marta Spinelli. Beyond power spectrum to unveil systematics on hi intensity maps. *arXiv preprint arXiv:2605.08895*, 2026.

[211] Pablo Lemos, Liam Parker, ChangHoon Hahn, Shirley Ho, Michael Eickenberg, Jiamin Hou, Elena Massara, Chirag Modi, Azadeh Moradinezhad Dizgah, Bruno Regaldo-Saint Blancard, et al. Simbig: Field-level simulation-based inference of galaxy clustering. *arXiv preprint arXiv:2310.15256*, 2023.

[212] Beatriz Tucci and Fabian Schmidt. EFTofLSS meets simulation-based inference: σ_8 from biased tracers. *JCAP*, 2024(5):063, May 2024.

[213] Oleg Savchenko, Florian List, Guillermo Franco Abellán, Noemi Anau Montel, and Christoph Weniger. Mean-field simulation-based inference for cosmological initial conditions. *arXiv preprint arXiv:2410.15808*, 2024.

[214] Oleg Savchenko, Guillermo Franco Abellán, Florian List, Noemi Anau Montel, and Christoph Weniger. Fast sampling of cosmological initial conditions with gaussian neural posterior estimation. *arXiv preprint arXiv:2502.03139*, 2025.

[215] Justine Zeghal, Denise Lanzieri, François Lanusse, Alexandre Boucaud, Gilles Louppe, Eric Aubourg, and Adrian E. Bayer. Simulation-based inference benchmark for weak lensing cosmology. *Astronomy & Astrophysics*, 699:A327, July 2025.

[216] Anirban Bairagi, Benjamin Wandelt, and Francisco Villaescusa-Navarro. How many simulations do we need for simulation-based inference in cosmology?, 2025.

[217] Adrian E. Bayer, Francisco Villaescusa-Navarro, Sammy Sharief, Romain Teyssier, Lehman H. Garrison, Laurence Perreault-Levasseur, Greg L. Bryan, Marco Gatti, and Eli Visbal. Field-level comparison and robustness analysis of cosmological n-body simulations. *The Astrophysical Journal*, 989(2):207, aug 2025.

[218] Anna Bonaldi, Ian Harrison, Stefano Camera, and Michael L. Brown. SKA weak lensing– II. Simulated performance and survey design considerations. *Monthly Notices of the Royal Astronomical Society*, 463(4):3686–3698, 08 2016.

[219] Stefano Camera, Ian Harrison, Anna Bonaldi, and Michael L. Brown. Ska weak lensing – iii. added value of multiwavelength synergies for the mitigation of systematics. *Monthly Notices of the Royal Astronomical Society*, 464(4):4747–4760, 10 2016.

[220] Héctor J Martínez, Manuel E Merchán, Carlos A Valotto, and Diego G Lambas. Quasar-galaxy and agn-galaxy cross-correlations. *The Astrophysical Journal*, 514(2):558, 1999.

[221] Bhuvnesh Jain, Ryan Scranton, and Ravi K. Sheth. Quasar—galaxy and galaxy—galaxy cross-correlations: model predictions with realistic galaxies. *Monthly Notices of the Royal Astronomical Society*, 345(1):62–70, 10 2003.

[222] Xiaohu Yang, H. J. Mo, Frank C. van den Bosch, Simone M. Weinmann, Cheng Li, and Y. P. Jing. The cross-correlation between galaxies and groups: probing the galaxy distribution in and around dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 362(2):711–726, 09 2005.

[223] Kerstin Paech, Nico Hamaus, Ben Hoyle, Matteo Costanzi, Tommaso Giannantonio, Steffen Hagstotz, Georg Sauerwein, and Jochen Weller. Cross-correlation of galaxies and galaxy clusters in the sloan digital sky survey and the importance of non-poissonian shot noise. *Monthly Notices of the Royal Astronomical Society*, 470(3):2566–2577, 06 2017.

[224] Michael R Nolta, EL Wright, L Page, CL Bennett, M Halpern, G Hinshaw, N Jarosik, A Kogut, M Limon, SS Meyer, et al. First year wilkinson microwave anisotropy probe observations: dark energy induced correlation with radio sources. *The Astrophysical Journal*, 608(1):10, 2004.

[225] Shirley Ho, Christopher Hirata, Nikhil Padmanabhan, Uros Seljak, and Neta Bahcall. Correlation of cmb with large-scale structure. i. integrated sachs-wolfe tomography and cosmological implications. *Physical Review D*, 78(4):043519, 2008.

[226] Christopher M Hirata, Shirley Ho, Nikhil Padmanabhan, Uroš Seljak, and Neta A Bahcall. Correlation of cmb with large-scale structure. ii. weak lensing. *Physical Review D*, 78(4):043520, 2008.

[227] A. Raccanelli, A. Bonaldi, M. Negrello, S. Matarrese, G. Tormen, and G. De Zotti. A reassessment of the evidence of the integrated sachs-wolfe effect through the wmap-nvss correlation. *Monthly Notices of the Royal Astronomical Society*, 386(4):2161–2166, 2008.

[228] Alvise Raccanelli, Gong-Bo Zhao, David J. Bacon, Matt J. Jarvis, Will J. Percival, Ray P. Norris, Huub Röttgering, Filipe B. Abdalla, Catherine M. Cress, Jean-Claude Kubwimana, Sam Lindsay, Robert C. Nichol, Mario G. Santos, and Dominik J. Schwarz. Cosmological measurements with forthcoming radio continuum surveys. *Monthly Notices of the Royal Astronomical Society*, 424(2):801–819, August 2012.

[229] Alvise Raccanelli, Olivier Doré, David J. Bacon, Roy Maartens, Mario G. Santos, Stefano Camera, Tamara M. Davis, Michael J. Drinkwater, Matt Jarvis, Ray Norris, and David Parkinson. Probing primordial non-Gaussianity via iSW measurements with SKA continuum surveys. *Journal of Cosmology and Astroparticle Physics*, 2015(1):042, January 2015.

[230] F. Bianchini et al. Cross-correlation between the CMB lensing potential measured by Planck and high-z sub-mm galaxies detected by the Herschel-ATLAS survey. *Astrophys. J.*, 802(1):64, 2015.

[231] Federico Bianchini and Andrea Lapi. Cross-correlation between cosmological and astrophysical datasets: the Planck and Herschel case. *IAU Symp.*, 306:202–205, 2014.

[232] Federico Bianchini et al. Toward a tomographic analysis of the cross-correlation between Planck CMB lensing and H-ATLAS galaxies. *Astrophys. J.*, 825(1):24, 2016.

[233] Suvodip Mukherjee, Benjamin D. Wandelt, and Joseph Silk. Multimessenger tests of gravity with weakly lensed gravitational waves. *Phys. Rev. D*, 101:103509, May 2020.

[234] Samuel J. Schmidt, Brice Ménard, Ryan Scranton, Christopher Morrison, and Cameron K. McBride. Recovering redshift distributions with cross-correlations: pushing the boundaries. *Monthly Notices of the Royal Astronomical Society*, 431(4):3307–3318, 04 2013.

[235] D. Alonso and P. G. Ferreira. Constraining ultralarge-scale cosmology with multiple tracers in optical and radio surveys. *Phys. Rev. D*, 92:063525, Sep 2015.

[236] Ely D. Kovetz, Alvise Raccanelli, and Mubdi Rahman. Cosmological constraints with clustering-based redshifts. *Monthly Notices of the Royal Astronomical Society*, 468(3):3650–3656, 03 2017.

[237] D. Alonso and P. G. Ferreira. Constraining ultralarge-scale cosmology with multiple tracers in optical and radio surveys. *Phys. Rev. D*, 92:063525, Sep 2015.

[238] L. Wolz, C. Tonini, C. Blake, and J. S. B. Wyithe. Intensity mapping cross-correlations: connecting the largest scales to galaxy evolution. *Monthly Notices of the Royal Astronomical Society*, 458(3):3399–3410, 03 2016.

[239] A. Pourtsidou, D. Bacon, R. Crittenden, and R. B. Metcalf. Prospects for clustering and lensing measurements with forthcoming intensity mapping and optical surveys. *Monthly Notices of the Royal Astronomical Society*, 459(1):863–870, 03 2016.

[240] Alkistis Pourtsidou. Testing gravity at large scales with HI intensity mapping. *Monthly Notices of the Royal Astronomical Society*, 461(2):1457–1464, 06 2016.

[241] Alvise Raccanelli, Ely Kovetz, Liang Dai, and Marc Kamionkowski. Detecting the integrated sachs-wolfe effect with high-redshift 21-cm surveys. *Phys. Rev. D*, 93:083512, Apr 2016.

[242] Alkistis Pourtsidou, David Bacon, and Robert Crittenden. HI and cosmological constraints from intensity mapping, optical and CMB surveys. *Monthly Notices of the Royal Astronomical Society*, 470(4):4251–4260, 06 2017.

[243] L. Wolz, C. Blake, and J. S. B. Wyithe. Determining the HI content of galaxies via intensity mapping cross-correlations. *Monthly Notices of the Royal Astronomical Society*, 470(3):3220–3226, 06 2017.

[244] David Alonso, Pedro G. Ferreira, Matt J. Jarvis, and Kavilan Moodley. Calibrating photometric redshifts with intensity mapping observations. *Phys. Rev. D*, 96:043515, Aug 2017.

[245] L Wolz, S G Murray, C Blake, and J S Wyithe. Intensity mapping cross-correlations II: HI halo models including shot noise. *Monthly Notices of the Royal Astronomical Society*, 484(1):1007–1020, 11 2018.

[246] Steven Cunnington, Ian Harrison, Alkistis Pourtsidou, and David Bacon. HI intensity mapping for clustering-based redshift estimation. *Monthly Notices of the Royal Astronomical Society*, 482(3):3341–3355, 10 2018.

[247] Ke Fang, Arka Banerjee, Eric Charles, and Yuuki Omori. A cross-correlation study of high-energy neutrinos and tracers of large-scale structure. *The Astrophysical Journal*, 894(2):112, 2020.

[248] Masamune Oguri. Measuring the distance-redshift relation with the cross-correlation of gravitational wave standard sirens and galaxies. *Physical Review D*, 93(8):083511, April 2016.

[249] Alvise Raccanelli, Ely D. Kovetz, Simeon Bird, Ilias Cholis, and Julian B. Muñoz. Determining the progenitors of merging black-hole binaries. *Phys. Rev. D*, 94:023516, Jul 2016.

[250] Giulio Scelfo, Nicola Bellomo, Alvise Raccanelli, Sabino Matarrese, and Licia Verde. GW×LSS: chasing the progenitors of merging binary black holes. *JCAP*, 2018(09):039–039, sep 2018.

[251] Giulio Scelfo, Lumen Boco, Andrea Lapi, and Matteo Viel. Exploring galaxies-gravitational waves cross-correlations as an astrophysical probe. *Journal of Cosmology and Astroparticle Physics*, 2020(10):045–045, oct 2020.

[252] Toshiya Namikawa, Atsushi Nishizawa, and Atsushi Taruya. Anisotropies of gravitational-wave standard sirens as a new cosmological probe without redshift information. *Physical review letters*, 116(12):121302, 2016.

[253] David Alonso, Giulia Cusin, Pedro G Ferreira, and Cyril Pitrou. Detecting the anisotropic astrophysical gravitational wave background in the presence of shot noise through cross-correlations. *Physical Review D*, 102(2):023002, 2020.

[254] Guadalupe Cañas Herrera, Omar Contigiani, and Valeri Vardanyan. Cross-correlation of the astrophysical gravitational-wave background with galaxy clustering. *Phys. Rev. D*, 102:043513, Aug 2020.

[255] Francesca Calore, Alessandro Cuoco, Tania Regimbau, Surabhi Sachdev, and Pasquale Dario Serpico. Cross-correlating galaxy catalogs and gravitational waves: a tomographic approach. *Physical Review Research*, 2(2):023314, 2020.

[256] Stefano Camera and Atsushi Nishizawa. Beyond concordance cosmology with magnification of gravitational-wave standard sirens. *Phys. Rev. Lett.*, 110:151103, Apr 2013.

[257] S. Libanore, M. C. Artale, D. Karagiannis, M. Liguori, N. Bartolo, Y. Bouffanais, N. Giacobbo, M. Mapelli, and S. Matarrese. Gravitational wave mergers as tracers of large scale structures. *Journal of Cosmology and Astroparticle Physics*, 2021(02):035–035, feb 2021.

[258] Suvodip Mukherjee, Benjamin D Wandelt, and Joseph Silk. Probing the theory of gravity with gravitational lensing of gravitational waves and galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 494(2):1956–1970, 03 2020.

[259] Suvodip Mukherjee, Benjamin D. Wandelt, Samaya M. Nissanke, and Alessandra Silvestri. Accurate precision cosmology with redshift unknown gravitational wave sources. *Phys. Rev. D*, 103:043520, Feb 2021.

[260] Suvodip Mukherjee and Joseph Silk. Time dependence of the astrophysical stochastic gravitational wave background. *Monthly Notices of the Royal Astronomical Society*, 491(4):4690–4701, 11 2019.

[261] Guadalupe Cañas-Herrera, Omar Contigiani, and Valeri Vardanyan. Learning how to surf: Reconstructing the propagation and origin of gravitational waves with gaussian processes. *The Astrophysical Journal*, 918(1):20, aug 2021.

[262] Cristina Cigarrán Díaz and Suvodip Mukherjee. Mapping the cosmic expansion history from LIGO-Virgo-KAGRA in synergy with DESI and SPHEREx. *Monthly Notices of the Royal Astronomical Society*, 511(2):2782–2795, 01 2022.

[263] Giulio Scelfo, Marta Spinelli, Alvise Raccanelli, Lumen Boco, Andrea Lapi, and Matteo Viel. Gravitational waves × HI intensity mapping: cosmological and astrophysical applications. *Journal of Cosmology and Astroparticle Physics*, 2022(01):004, jan 2022.

[264] Giulio Scelfo, Maria Berti, Alessandra Silvestri, and Matteo Viel. Testing gravity with gravitational waves × electromagnetic probes cross-correlations. *Journal of Cosmology and Astroparticle Physics*, 2023(02):010, feb 2023.

[265] M. Bosi, N. Bellomo, and A. Raccanelli. Constraining extended cosmologies with gw×lss cross-correlations. *Journal of Cosmology and Astroparticle Physics*, 2023(11):086, nov 2023.

[266] Federico Semenzato, J Andrew Casey-Clyde, Chiara MF Mingarelli, Alvise Raccanelli, Nicola Bellomo, Nicola Bartolo, and Daniele Bertacca. Cross-correlating the universe: The gravitational wave background and large-scale structure. *arXiv preprint arXiv:2411.00532*, 2024.

[267] Francisco Villaescusa-Navarro, Daniel Anglés-Alcázar, Shy Genel, David N Spergel, Yin Li, Benjamin Wandelt, Andrina Nicola, Leander Thiele, Sultan Hassan, Jose Manuel Zorrilla Matilla, et al. Multifield cosmology with artificial intelligence. *arXiv preprint arXiv:2109.09747*, 2021.

[268] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[269] Martín Crocce, Sebastián Pueblas, and Román Scoccimarro. Transients from initial conditions in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 373(1):369–381, 10 2006.

[270] Francisco Villaescusa-Navarro, Shy Genel, Emanuele Castorina, Andrej Obuljen, David N. Spergel, Lars Hernquist, Dylan Nelson, Isabella P. Carucci, Annalisa Pillepich, Federico Marinacci, Benedikt Diemer, Mark Vogelsberger, Rainer Weinberger, and Rüdiger Pakmor. Ingredients for 21 cm Intensity Mapping. *The Astrophysical Journal*, 866(2):135, October 2018. Publisher: The American Astronomical Society.

[271] Francisco Villaescusa-Navarro, Shy Genel, Daniel Angles-Alcazar, Leander Thiele, Romeel Dave, Desika Narayanan, Andrina Nicola, Yin Li, Pablo Villanueva-Domingo, Benjamin Wandelt, David N. Spergel, Rachel S. Somerville, Jose Manuel Zorrilla Matilla, Faizan G. Mohammad, Sultan Hassan, Helen Shao, Digvijay Wadekar, Michael Eickenberg, Kaze W. K. Wong, Gabriella Contardo, Yongseok Jo, Emily Moser, Erwin T. Lau, Luis Fernando Machado Poletti Valle, Lucia A. Perez, Daisuke Nagai, Nicholas Battaglia, and Mark Vogelsberger. The CAMELS Multifield Dataset: Learning the Universe's Fundamental Parameters with Artificial Intelligence. *The Astrophysical Journal Supplement Series*, 259(2):61, April 2022. arXiv:2109.10915 [astro-ph].

[272] George Papamakarios and Iain Murray. Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. *arXiv e-prints*, page arXiv:1605.06376, May 2016.

[273] Jan-Matthis Lueckmann, Pedro J. Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H. Macke. Flexible statistical inference for mechanistic models of neural dynamics. *arXiv e-prints*, page arXiv:1711.01861, November 2017.

[274] David S. Greenberg, Marcel Nonnenmacher, and Jakob H. Macke. Automatic Posterior Transformation for Likelihood-Free Inference. *arXiv e-prints*, page arXiv:1905.07488, May 2019.

[275] Michael Deistler, Pedro J Goncalves, and Jakob H Macke. Truncated proposals for scalable and hassle-free simulation-based inference. *arXiv e-prints*, page arXiv:2210.04815, October 2022.

[276] Ivan Kobyzev, Simon J.D. Prince, and Marcus A. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3964–3979, 2021.

[277] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. *Advances in neural information processing systems*, 30, 2017.

[278] Jan Boelts, Michael Deistler, Manuel Gloeckler, Álvaro Tejero-Cantero, Jan-Matthis Lueckmann, Guy Moss, Peter Steinbach, Thomas Moreau, Fabio Muratore, Julia Linhart, et al. sbi reloaded: a toolkit for simulation-based inference workflows. *arXiv preprint arXiv:2411.17337*, 2024.

[279] Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer, 2017.

[280] G. Autieri, M. Berti, M. Spinelli, B. S. Haridasu, and M. Viel. Weighing neutrinos with 21cm intensity mapping at the SKAO. *JCAP*, 2026(1):050, January 2026.

[281] I. P. Carucci, J. L. Bernal, S. Cunnington, M. G. Santos, J. Wang, J. Fonseca, K. Grainge, M. O. Irfan, Y. Li, A. Pourtsidou, M. Spinelli, and L. Wolz. Hydrogen intensity mapping with meerkat: Preserving cosmological signal by optimising contaminant separation. *A&A*, 703:A222, November 2025.

[282] Maria Berti, Marta Spinelli, and Matteo Viel. 21 cm intensity mapping cross-correlation with galaxy surveys: Current and forecasted cosmological parameters estimation for the SKAO. *MNRAS*, 529(4):4803–4817, April 2024.

[283] YP Jing. Correcting for the alias effect when measuring the power spectrum using a fast fourier transform. *The Astrophysical Journal*, 620(2):559–563, 2005.