Generating Mock Galaxy Catalogs

Carl Audric Guia[1]

[1]*Harvard University, Cambridge, MA 02138, USA*

Submitted to Physics 212 at Harvard University

ABSTRACT

Mock galaxy catalogs are important in understanding the large scale structure of the universe and forecasting cosmological surveys. In this project, mock catalogs are constructed via forward modeling that begins with the linear power spectra generated by `CLASS`. The pipeline progresses through Gaussian field generation, bias expansion with stochastic noise, and perturbative predictions from `CLASS-PT`. Two cosmological models are considered: Planck ΛCDM and a $w_0 w_a$ dark energy model motivated by DESI DR2 forecasts. The result is then compared with theoretical predictions from `CLASS-PT` to ensure consistency. In the perturbative framework, the nonlinear corrections from $P_{\text{1-loop, SPT}}$ are disabled. The simulated and theoretical power spectra of the galaxy fields exhibit consistent trends in redshift evolution and amplitude. In the $w_0 w_a$ model, suppression at low k is observed across simulation and theory. This validates the forward-modeling pipeline of mock galaxy catalogs and supports its use in future work. More quantitative comparisons with perturbation theory and cosmological inference via `MCMC` are recommended.

Keywords: Large-scale structure of the universe (902) — Cosmological perturbation theory (341) — Density parameters (372) — Galaxy formation (595)

1. INTRODUCTION

The large-scale distribution of galaxies is central to cosmology. Galaxy surveys provide a window into the underlying matter field, and since galaxies are biased tracers of this structure, their distribution is shaped by a complex interplay of gravitational evolution and cosmological parameters. In bridging the gap between the initial matter field and the observed galaxy density today, mock galaxy catalogs play a critical role. They enable detailed tests of bias models, shot noise, and cosmic variance under controlled conditions, and they can validate pipelines in cosmological inference.

This project constructs mock galaxy catalogs using a forward-modeling approach based on field-level perturbation theory. Two cosmological models are considered: Planck ΛCDM (Planck Collaboration 2020) and a $w_0 w_a$ dark energy model motivated by DESI DR2 forecasts (DESI Collaboration 2025). To isolate the potential effects of evolving dark energy on structure formation, the background cosmological parameters are fixed to the Planck 2018 ΛCDM best-fit values for both models. This way, any differences in the power spectra are solely from the time-dependent dark energy equation of state in the $w_0 w_a$ model.

Starting from linear power spectra generated using `CLASS` (D. Blas et al. 2011), Gaussian fields are sampled on a grid, and bias expansion is implemented to generate galaxy fields. The bias expansion includes linear, quadratic, and tidal terms, as well as a stochastic noise component (M. Schmittfull 2019). For both cosmological models, the tracer number density is fixed at $\bar{n} = 10^{-3} \, h^3/\text{Mpc}^3$, which is the constraint for the Bright Galaxy Survey (BGS) as found by the DESI Collaboration (2025) forecasts. Similarly consistent with DESI Collaboration (2016) and S. Chen et al. (2021) are the chosen bias parameters $b_1 = 1.2$, $b_2 = -0.405$, and $b_{G_2} = -0.127$.

After computing the power spectra of the mock galaxy fields, the results are compared across redshift and cosmological models, as well as against theoretical predictions from `CLASS-PT` (A. Chudaykin et al. 2020), a perturbative Boltzmann code that includes 1-loop corrections. For consistency, the comparison is restricted to non-loop bias.

Email: cgguia@college.harvard.edu

2. METHODS

2.1. *Overview of the Simulation Pipeline*

The pipeline automates the process of generating simulation outputs from theory using a sequence of Python scripts. It includes the following stages:

1. Defining the fiducial cosmologies in the YAML files grounded in Planck 2018 and DESI 2025 forecasts.

2. Generating linear matter power spectra at multiple redshifts using the `CLASS` Boltzmann solver.

3. Constructing Gaussian initial conditions on a 3D grid by sampling $P(k)$ via FFT.

4. Applying galaxy bias expansion to generate mock galaxy fields with shot noise.

5. Computing power spectra for the Gaussian and galaxy fields binned in logarithmic k-space.

6. Generating `CLASS-PT` predictions to validate simulation outputs.

2.2. *Power Spectrum Generation with CLASS*

To compute the linear matter power spectrum $P(k)$, the public Boltzmann solver `CLASS` (D. Blas et al. 2011) was used. It accepts cosmological parameters defined via YAML. For the Planck ΛCDM model, the parameters $H_0 = 67.36$, $\Omega_b h^2 = 0.02237$, $\Omega_c h^2 = 0.1200$ and one massive neutrino species ($N_{\mathrm{ncdm}} = 1$, $m_{\mathrm{ncdm}} = 0.06$) were taken from the Planck Collaboration (2020). For the $w_0 w_a$ model, the same background parameters were used to isolate the impact of dynamical dark energy on structure growth except for $w_0 = -0.838$, $w_a = -0.62$, and $\Omega_{\mathrm{fld}} = 0.6847$ constrained by the DESI Collaboration (2025).

To ensure numerical stability in the phantom regime, the Parameterized Post-Friedmann (PPF) framework was activated for the dynamical dark energy model. CLASS outputs $P(k)$ at redshifts $z = 0, 0.5, 1, 2, 3$ as two-column text files. These spectra serve as the input for generating the Gaussian field.

2.3. *Generating the Gaussian Field*

The Gaussian field is sampled in Fourier space on a 256^3 grid in a $1000\,h^{-1}$Mpc box. Each k-mode is drawn as a complex Gaussian variable scaled by:

$$\delta(\vec{k}) = [\mathcal{N}(0,1) + i\mathcal{N}(0,1)] \cdot \sqrt{\frac{P(k)}{2}} \tag{1}$$

where $P(k) = \langle |\delta(k)|^2 \rangle = 2\sigma^2$. An inverse FFT with forward normalization brings the field into real space. The result is a mean-zero field with variance changing with the evolving redshift.

2.4. *Bias Expansion and Galaxy Field Construction*

To convert the Gaussian field into a mock galaxy field, the bias expansion by M. Schmittfull (2019) is adopted:

$$\delta_h(\vec{x}) = b_1 \tilde{\delta}_1 + b_2(\tilde{\delta}_1^2 - \langle \tilde{\delta}_1^2 \rangle) + b_{G_2} \tilde{G}_2 + \epsilon \tag{2}$$

Here, $\tilde{\delta}_1(\vec{x}) = \delta_1(\vec{q} + \vec{\psi}_1(\vec{q}))$ is the displaced Gaussian field via the Zel'dovich approximation at first order. $\tilde{G}_2$, the tidal field operator, is implemented as follows:

$$G_2(\vec{q}) \equiv \left[\frac{\partial_i \partial_j}{\partial^2} \delta_1(\vec{q}) \right]^2 - \delta_1^2(\vec{q}) \tag{3}$$

Lastly, $\epsilon(\vec{x}) \sim \mathcal{N}\left(0, \frac{1}{\bar{n}V}\right)$ is the shot noise which models stochasticity. V is the volume of the grid cell for proper scaling, and the mean number density $\bar{n} = 10^{-3}\,h^3/\mathrm{Mpc}^3$ is chosen consistent with the DESI Collaboration (2025) BGS samples. The shot noise adds a flat noise component to the galaxy power spectrum. The operators are shifted to Eulerian space using cloud-in-cell interpolation, which is also used by M. Schmittfull (2019).

The bias parameters are fixed across redshifts and cosmologies as:

$$b_1 = 1.2, \quad b_2 = -0.405, \quad b_{G_2} = -0.127 \tag{4}$$

These are motivated by DESI Collaboration (2016) BGS forecasts and simulation-based fits by S. Chen et al. (2021).

2.5. *Power Spectrum Computation and Binning*

To compute the power spectrum of the Gaussian and galaxy fields, FFT of the real-space density field was performed using `numpy.fft.fftn`, again with forward normalization to ensure consistency with the input power spectrum. The resulting Fourier-space grid includes modes up to the Nyquist frequency, which is defined as

$$k_{\mathrm{Nyq}} = \frac{\pi}{\Delta x} = \frac{\pi \cdot n_{\mathrm{grid}}}{L}$$

k_{Nyq} represents the smallest wavelength that can be unambiguously sampled on the discrete grid, a foundational result of the Shannon Sampling Theorem (C. E. Shannon 1949).

The power spectrum is computed from the squared Fourier amplitudes and then binned into logarithmically spaced k-bins up to k_{Nyq}. This ensures that all modes contributing to the power spectrum are below the aliasing threshold, and the resulting $P(k)$ is not undersampled. Up to a factor of 2π, the power spectrum is

$$P(k) = |\delta_k|^2 \cdot V \tag{5}$$

The Planck ΛCDM universe is binned into 8 intervals, while the $w_0 w_a$ model is binned into 7 to balance resolution with statistical noise at low and high k.

2.6. *CLASS-PT Prediction and Parameter Setup*

`CLASS-PT` (A. Chudaykin et al. 2020) was used to generate theoretical galaxy power spectra. It is a perturbative extension of `CLASS` that includes 1-loop perturbation theory and bias modeling. For consistency with the mock galaxy field pipeline, non-linear corrections are excluded by setting:

$$cs_0 = cs_2 = cs_4 = b_{\Gamma_3} = b_4 = 0 \tag{6}$$

This removes $P_{1-\mathrm{loop,SPT}}$ from the total spectrum as described in Equation (2.10) of A. Chudaykin et al. (2020). The bias parameters $b_1, b_2,$ and b_{G2} in the mock galaxy field are retained to allow for direct $P_{gg}(k)$ comparisons with CLASS-PT outputs across all redshifts and cosmologies.
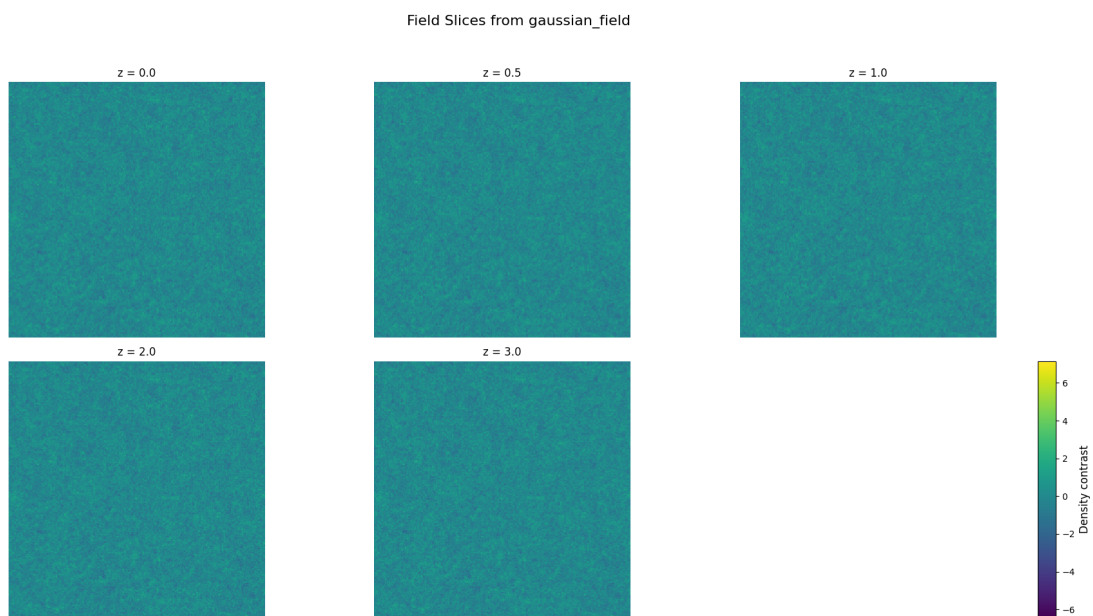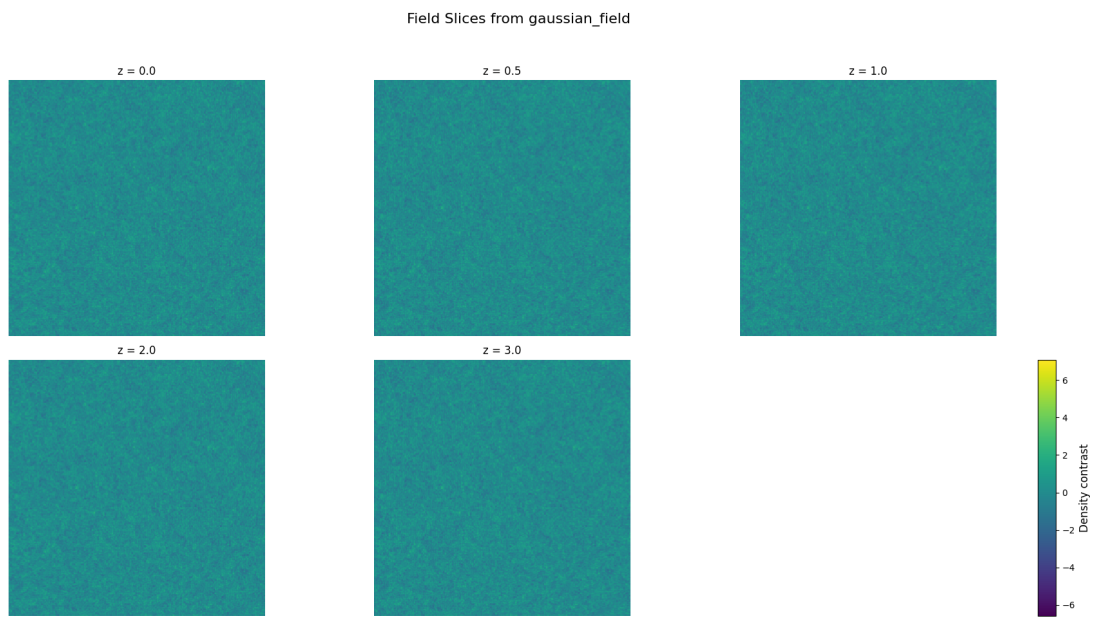
3. RESULTS

3.1. *Field-Level Statistics*

The variance of the Gaussian fields decreases with redshift due to suppressed growth of structure at early times. The galaxy fields, on the other hand, exhibit higher variance across all redshifts as shown in Table 1. This amplification is consistent with the inclusion of linear and nonlinear bias terms in the expansion. Figure 2 also aligns with this amplification by showing the enhanced density contrast of the galaxy fields. The variance remains relatively flat from $z = 0$ to $z = 3$, reflecting the dominant contribution of biasing and shot noise, but the trend is still decreasing as anticipated. Also, the variances between the two models are very similar, an observation supported by the nearly identical power spectra below.

Table 1. Variance of Gaussian and galaxy fields across redshift for Planck ΛCDM and $w_0 w_a$ cosmological models. Each value corresponds to the variance of the field computed on a 256^3 grid in a $1000\,h^{-1}$Mpc box.

Model	Field	$z = 0.0$	$z = 0.5$	$z = 1.0$	$z = 2.0$	$z = 3.0$
Planck ΛCDM	Gaussian	2.40101	1.42006	0.88660	0.42030	0.24007
Planck ΛCDM	Galaxy	22.07993	19.44727	18.29369	17.44254	17.14807
$w_0 w_a$	Gaussian	2.42002	1.44392	0.90267	0.42381	0.24199
$w_0 w_a$	Galaxy	22.11306	19.50880	18.33609	17.43391	17.14452

Field Slices from gaussian_field

z = 0.0   z = 0.5   z = 1.0

z = 2.0   z = 3.0

Density contrast

(a) Planck ΛCDM

Field Slices from gaussian_field

z = 0.0   z = 0.5   z = 1.0

z = 2.0   z = 3.0

Density contrast

(b) $w_0 w_a$ model

Figure 1. Slices of the real-space Gaussian field $\delta(x)$ for both cosmological models. Each field is generated from the linear power spectrum using a common grid and random seed.

Field Slices from galaxy_field

(a) Planck ΛCDM

Field Slices from galaxy_field
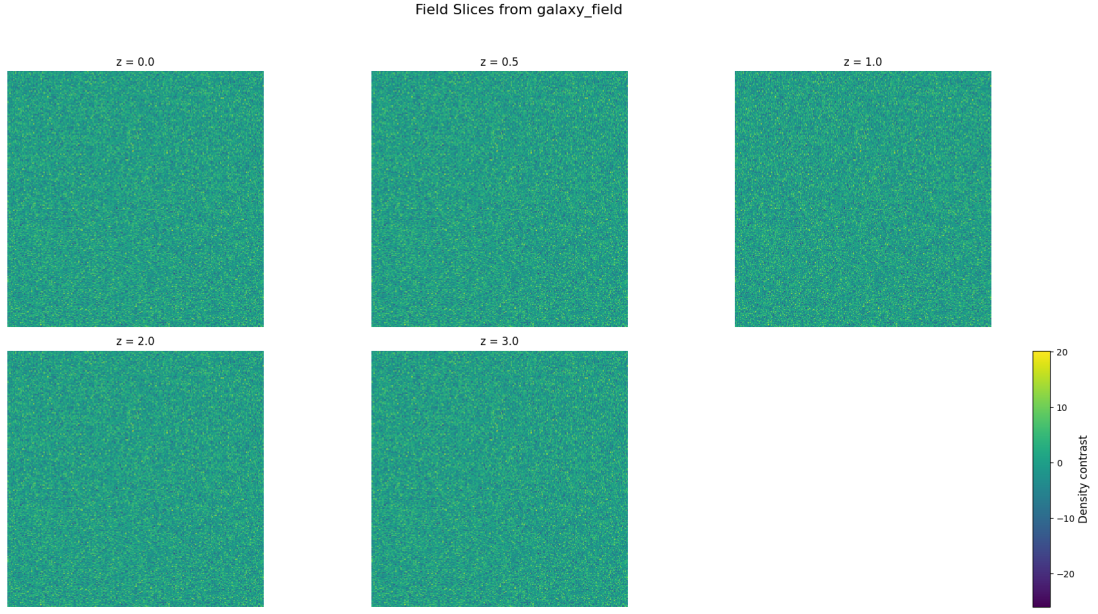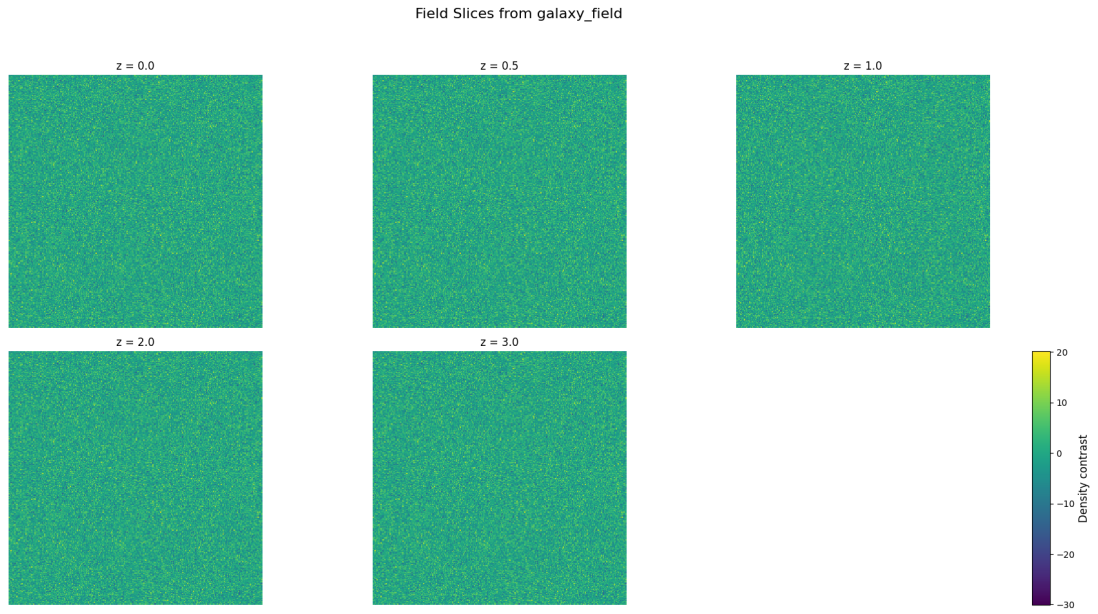
(b) $w_0 w_a$ model

Figure 2. Slices of the real-space galaxy field $\delta_h(x)$ for both cosmological models, generated by applying the bias expansion and shot noise to the underlying Gaussian matter field.
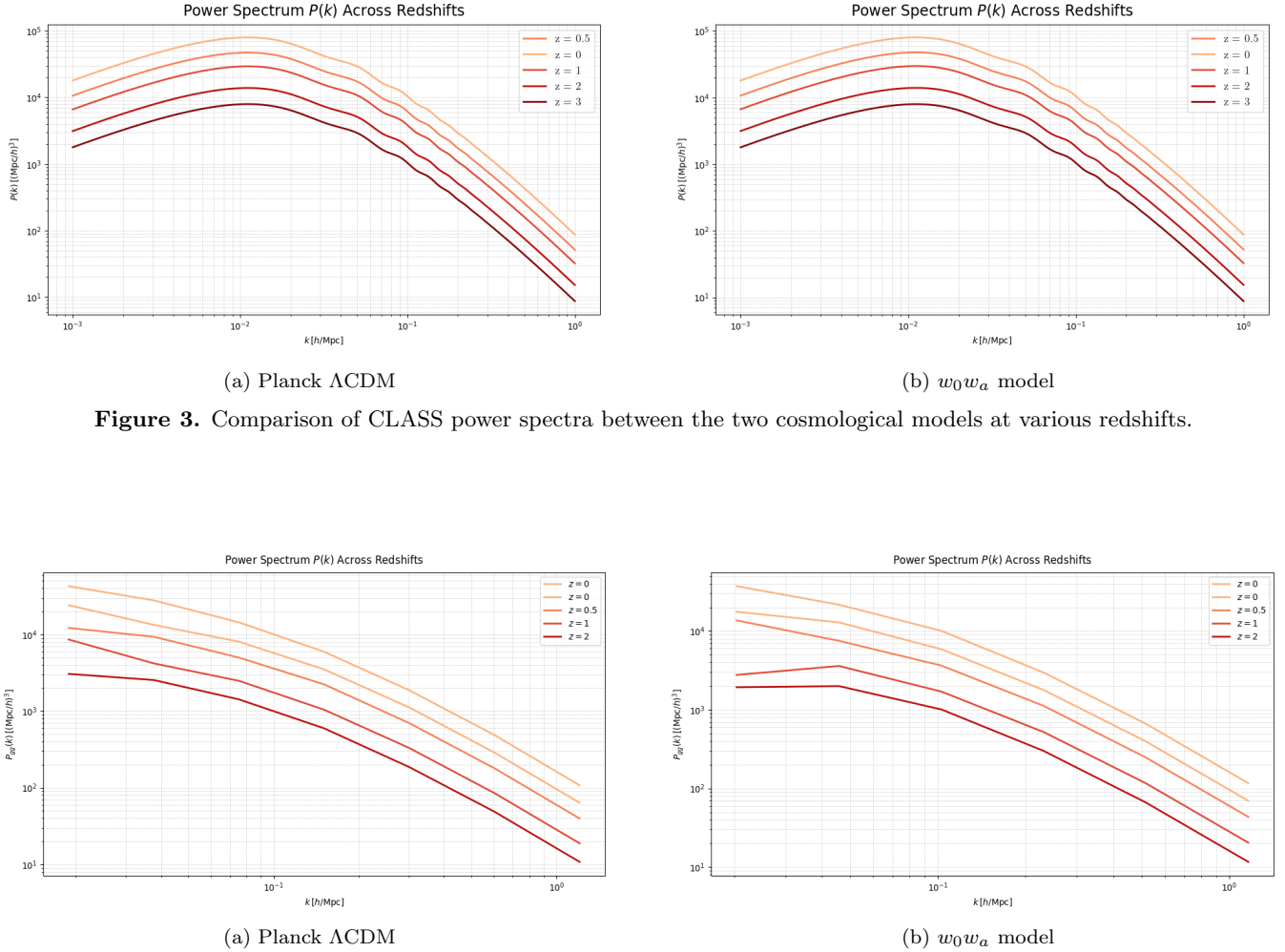
3.2. *Mock Galaxy Power Spectra*

The CLASS-computed power spectra in Figure 3 show a consistent decline in amplitude with increasing redshift, reflecting the suppression of structure growth in early times. Small-amplitude oscillations are visible in the intermediate k-range, corresponding to baryon acoustic oscillations (BAOs) imprinted at recombination. The linear power spectra produced by CLASS are nearly indistinguishable between the two cosmologies across redshifts. This reflects the fact that early-time physics is identical between the two models. This result might hint at the insensitivity of the linear power spectrum to the dark energy equation of state.

Figure 4 displays the power spectra of the Gaussian density fields for the two cosmologies. As expected, the amplitude of $P(k)$ declines with increasing redshift. The spectra from both models exhibit a similar shape consistent with Figure 3. The BAO features appear somewhat jagged due to limited k-resolution from binning: with only 7–8 logarithmic bins used to reduce noise, individual BAO wiggles are not fully resolved.

Figure 5 shows the galaxy power spectra derived from the bias expansion applied to the Gaussian fields, incorporating both nonlinear bias terms and stochastic shot noise. Compared to the Gaussian field spectra, the galaxy spectra exhibit an overall enhancement in amplitude particularly at low k. This might be due to the contribution of the linear bias term b_1^2. A flat high-k tail is visible in both cosmologies as expected from the stochastic noise $P_\epsilon(k) \sim 1/\bar{n}$ from $\epsilon(x)$. The amplitude of this tail matches $\bar{n} = 10^{-3}\,h^3/\mathrm{Mpc}^3$. These results validate the implementation of the bias expansion including stochastic noise.

(a) Planck ΛCDM   (b) $w_0 w_a$ model

Figure 3. Comparison of CLASS power spectra between the two cosmological models at various redshifts.

(a) Planck ΛCDM   (b) $w_0 w_a$ model

Figure 4. Comparison of Gaussian field power spectra between the two cosmological models at various redshifts.
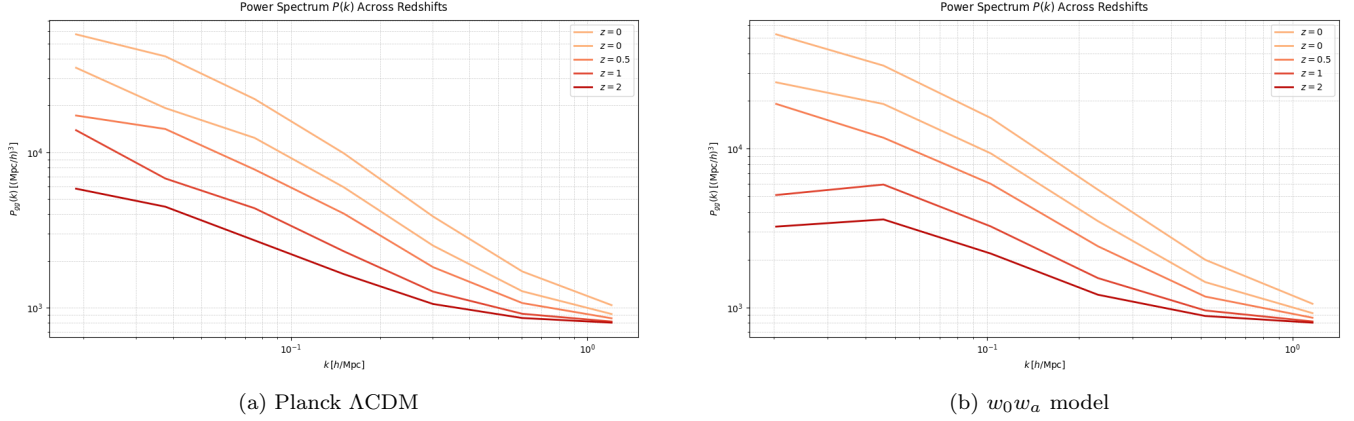
Figure 5. Galaxy power spectra with biasing and shot noise for Planck ΛCDM and $w_0 w_a$ cosmologies across redshift.

3.3. *Comparison with CLASS-PT*

Figure 6 shows the galaxy power spectra predicted by `CLASS-PT` for both cosmologies. With non-loop bias only, the overall shape and redshift evolution of these spectra resemble those of the mock galaxy fields in Figure 5 in the applicable k-domain. As in the mock fields, the amplitude decreases with redshift. A flat high-k tail is present as well, something that was not observed with only `CLASS`. The similarity in structure and redshift trends between the simulations and `CLASS-PT` validates the consistency of the field-level bias expansion.
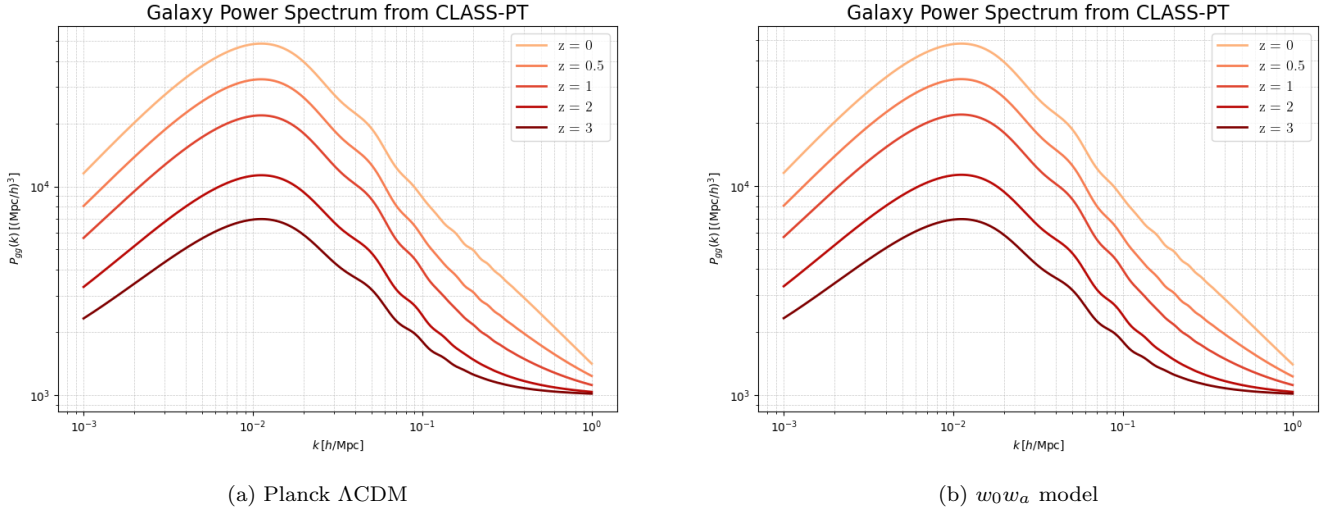
Figure 6. CLASS-PT predictions for the galaxy power spectrum with non-loop bias for both cosmological models.

4. RECOMMENDATIONS

While the results show qualitative agreement with `CLASS-PT` predictions and highlight the effects of bias in structure growth, several natural extensions would bolster the robustness of this work.

4.1. *Quantitative Comparison with Perturbation Theory*

Future work could benefit from a more quantitative analysis. This includes computing the residuals and RMS fractional errors between the simulated and `CLASS-PT` power spectra across redshifts and scales. Such a comparison would clarify the parameter space where non-loop perturbation theory remains valid.

4.2. *Likelihood Construction and Cosmological Inference*

The outputs of the simulation can be treated as synthetic observations and compared to theoretical predictions from `CLASS-PT` using an `MCMC` sampler such as `Cobaya` or `emcee`. This would provide a consistency check: by treating b_1, b_2, b_{G_2} and c_s^2 as free parameters, one could test whether the input parameters can be recovered to generate the mock galaxy fields.

In principle, c_s^2 should be included in this analysis, even though it was set to zero in the present work. It accounts for small-scale nonlinear corrections from the nonlinear matter density and is related to the resolution of the simulated field. While it is justified to set $c_s^2 = 0$ when modeling only non-loop bias on linear Gaussian field, a more complete framework should account for this parameter to minimize systematic errors.

4.3. *Higher-Order Statistics*

The bispectrum and other higher-order statistics could be computed from the simulated fields. This would test whether the nonlinear bias model captures three-point correlations, which can be a probe of primordial non-Gaussianity f_{NL}. Computing the bispectrum would also provide more insight into the effects of massive neutrinos at different scales.

ACKNOWLEDGMENTS

REFERENCES

Blas, D., L. J., & Tram, T. 2011, Journal of Cosmology and Astroparticle Physics, 2011, 1, doi: 10.1088/1475-7516/2011/07/034

Chen, S., Lee, H., & Dvorkin, C. 2021, Journal of Cosmology and Astroparticle Physics, 2021, 1, doi: 10.1088/1475-7516/2021/11/037

Chudaykin, A., Ivanov, M. M., Philcox, O. H. E., & Simonović, M. 2020, Phys. Rev. D, 102, 063533, doi: 10.1103/PhysRevD.102.063533

Collaboration, D. 2016, arXiv e-prints. https://arxiv.org/abs/1611.00036

Collaboration, D. 2025, Journal of Cosmology and Astroparticle Physics, 02, 1, doi: 10.1088/1475-7516/2025/02/021

Collaboration, P. 2020, Astronomy & Astrophysics, 641, 1, doi: 10.1051/0004-6361/201833910

Schmittfull, M. et al. 2019, Physical Review D, 100, 1, doi: 10.1103/PhysRevD.100.043514

Shannon, C. E. 1949, Proceedings of the I.R.E., 37, 10, doi: 10.1109/JRPROC.1949.232969