*Astronomy & Astrophysics* manuscript no. forestflow
July 25, 2024


# ForestFlow: cosmological emulation of Lyman-$\alpha$ forest clustering from linear to nonlinear scales

J. Chaves-Montero[1⋆], L. Cabayol-Garcia[1,2⋆⋆], M. Lokken[1], A. Font-Ribera[1⋆⋆⋆], and others

[1] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, 08193 Bellaterra (Barcelona), Spain
[2] Port d'Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain

Received 2024; accepted XXX

**ABSTRACT**

On large scales, measurements of the Lyman-$\alpha$ forest offer insights into the expansion history of the Universe, while on small scales, these impose strict constraints on the growth history, the nature of dark matter, and the sum of neutrino masses. This work introduces FORESTFLOW, a cosmological emulator designed to bridge the gap between large- and small-scale analyses. Using conditional normalizing flows, FORESTFLOW emulates the 2 Lyman-$\alpha$ linear biases ($b_\delta$ and $b_\eta$) and 6 parameters describing small-scale deviations of the three-dimensional flux power spectrum ($P_{\rm 3D}$) from linear theory. These 8 parameters are modeled as a function of cosmology — the small-scale amplitude and slope of the linear power spectrum —- and the physics of the intergalactic medium. Thus, in combination with a Boltzmann solver, FORESTFLOW can predict $P_{\rm 3D}$ on arbitrarily large (linear) scales and the one-dimensional flux power spectrum ($P_{\rm 1D}$) — the primary observable for small-scale analyses — without the need for interpolation or extrapolation. Trained on a suite of 30 fixed-and-paired cosmological hydrodynamical simulations spanning redshifts from $z = 2$ to 4.5, FORESTFLOW achieves 3 and 1.5% precision in describing $P_{\rm 3D}$ and $P_{\rm 1D}$ from linear scales to $k = 5\,{\rm Mpc}^{-1}$ and $k_\parallel = 4\,{\rm Mpc}^{-1}$, respectively. Thanks to its parameterization, the precision of the emulator is similar for two extensions to the $\Lambda$CDM model — massive neutrinos and curvature — and ionization histories not included in the training set. FORESTFLOW will be crucial for the cosmological analysis of Lyman-$\alpha$ forest measurements from the DESI survey and facilitate novel multiscale analyses.

**Key words.** quasars: absorption lines – cosmology: large-scale structure of Universe – methods: statistical

## 1. Introduction

The Lyman-$\alpha$ forest refers to absorption lines in the spectra of high-redshift quasars resulting from Lyman-$\alpha$ absorption by neutral hydrogen in the intergalactic medium (IGM; for a review, see McQuinn 2016). Even though quasars can be observed at very high redshifts with relatively short exposure times, the scarcity of these sources limits their direct use for precision cosmology. Conversely, Lyman-$\alpha$ forest measurements from a single quasar spectrum provide information about density fluctuations over hundreds of megaparsecs along the line of sight, making this observable an excellent tracer of large-scale structure at high redshifts.

Cosmological analyses of the Lyman-$\alpha$ forest rely on either three-dimensional correlations of the Lyman-$\alpha$ transmission field ($\xi_{\rm 3D}$; e.g.; Slosar et al. 2011) or correlations along the line-of-sight of each quasar; i.e., the one-dimensional flux power spectrum ($P_{\rm 1D}$; e.g.; Croft et al. 1998; McDonald et al. 2000). The first analyses set constraints on the expansion history of the Universe by measuring baryonic acoustic oscillations (BAO; e.g.; Busca et al. 2013; Slosar et al. 2013; du Mas des Bourboux et al. 2020), for which linear or perturbation theory is accurate enough. On the other hand, $P_{\rm 1D}$ analyses measure the small-scale amplitude and slope of the linear power spectrum (e.g.; Croft et al. 1998; McDonald et al. 2000; Zaldarriaga et al. 2001; Viel et al. 2004; McDonald et al. 2005), the nature of dark matter (e.g.; Seljak et al. 2006; Viel et al. 2013; Iršič et al. 2017; Palanque-Delabrouille et al. 2020; Rogers & Peiris 2021b; Iršič et al. 2024), the thermal history of the IGM (e.g.; Viel & Haehnelt 2006; Bolton et al. 2008; Lee et al. 2015; Walther et al. 2019; Boera et al. 2019; Gaikwad et al. 2020, 2021) and the reionization history of the Universe (see the reviews Meiksin 2009; McQuinn 2016). In combination with cosmic microwave background constraints, $P_{\rm 1D}$ analyses also set tight constraints on the sum of neutrino masses and the running of the spectral index (e.g.; Spergel et al. 2003; Verde et al. 2003; Viel et al. 2004; Seljak et al. 2005, 2006; Palanque-Delabrouille et al. 2015, 2020).

Unlike $\xi_{\rm 3D}$ studies, $P_{\rm 1D}$ analyses go deep into the nonlinear regime and require time-demanding hydrodynamical simulations (e.g.; Cen et al. 1994; Miralda-Escudé et al. 1996; Meiksin et al. 2001; Lukić et al. 2015; Bolton et al. 2017; Walther et al. 2021; Chabanier et al. 2023; Puchwein et al. 2023; Bird et al. 2023). Naive analyses would demand running millions of hydrodynamical simulations, which is currently unfeasible. Rather, the preferred solution is constructing fast surrogate models that make precise predictions across the input parameter space using simulation measurements as the training set. The main advantage of these surrogate models, known as emulators, is reducing the number of simulations required for Bayesian inference from millions to dozens or hundreds. In the context of Lyman-$\alpha$ forest studies, the first $P_{\rm 1D}$ emulators involved simple linear interpolation (McDonald et al. 2006) and progressively moved towards using Gaussian processes (GPs; Sacks et al. 1989; MacKay et al. 1998) and neural networks (NNs; McCulloch & Pitts 1943); for instance, Bird et al. (2019); Rogers et al. (2019); Walther et al.

⋆ jchaves@ifae.es
⋆⋆ lcabayol@pic.es
⋆⋆⋆ afont@ifae.es

(2019); Pedersen et al. (2021); Takhtaganov et al. (2021); Rogers & Peiris (2021a); Fernandez et al. (2022); Bird et al. (2023); Molaro et al. (2023); Cabayol-Garcia et al. (2023).

The primary purpose of this work is to provide consistent predictions for Lyman-$\alpha$ forest clustering from linear to nonlinear scales. There are three main approaches to achieve this. The first relies on perturbation theory (e.g.; Givans & Hirata 2020; Chen et al. 2021; Ivanov 2024), which delivers precise predictions on perturbative scales at the cost of marginalizing over a large number of free parameters. The second involves emulating power spectrum modes measured from a suite of cosmological hydrodynamical simulations, which provides precise predictions from quasilinear to nonlinear scales. The main limitation of this approach is that accessing the largest scales used in BAO analyses, $r \simeq 300$ Mpc, would require hydrodynamical simulations at least 3 times larger than this scale (e.g.; Angulo et al. 2008), which is currently unfeasible due to the computational demands of these simulations.

Instead, we follow the third approach of emulating the best-fitting parameters of a physically-motivated Lyman-$\alpha$ clustering model to measurements from a suite of cosmological hydrodynamical simulations (see McDonald 2003; Arinyo-i-Prats et al. 2015). We emulate the 2 Lyman-$\alpha$ linear biases ($b_\delta$ and $b_\eta$), which completely set the large-scale behavior of $P_{3D}$ together with the linear power spectrum, and 6 parameters modeling small-scale deviations of $P_{3D}$ from linear theory. Consequently, this strategy has the potential to make precise $P_{3D}$ predictions from nonlinear to arbitrarily large (linear) scales even when using simulations with moderate sizes as training data. It also enables predicting any Lyman-$\alpha$ statistic derived from $P_{3D}$ without requiring interpolation or extrapolation. For instance, we can compute $\xi_{3D}$ by taking the Fourier transform of $P_{3D}$ or determine $P_{1D}$ by integrating its perpendicular modes

$$P_{1D}(k_\parallel) = (2\pi)^{-1} \int_0^\infty \mathrm{d}k_\perp \, k_\perp \, P_{3D}(k_\parallel, k_\perp), \quad (1)$$

where $k_\parallel$ and $k_\perp$ indicate parallel and perpendicular modes, respectively.

We emulate the 8 previous parameters as a function of cosmology and IGM physics using FORESTFLOW[1], a conditional normalizing flow (cNFs; Winkler et al. 2019; Papamakarios et al. 2019). FORESTFLOW predicts the values and correlations of model parameters, allowing the propagation of these correlations onto $P_{3D}$ and derived statistics. We train FORESTFLOW using measurements from the suite of cosmological hydrodynamical simulations presented in Pedersen et al. (2021), which consists of 30 fixed-and-paired hydrodynamical simulations of 67.5 Mpc on a side.

The release of FORESTFLOW is quite timely for BAO and $P_{1D}$ analyses of the ongoing Dark Energy Spectroscopic Instrument survey (DESI; DESI Collaboration et al. 2016), which will quadruple the number of line-of-sights observed by first the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) and its extension (eBOSS; Dawson et al. 2016). DESI has already proven the constraining power of Lyman-$\alpha$ studies by measuring the isotropic BAO scale with $\simeq 1\%$ precision from the Data Release 1 (DESI Collaboration et al. 2024) and $P_{1D}$ at 9 redshift bins with a precision of a few percent from the Early Data Release (Ravoux et al. 2023; Karaçaylı et al. 2024). In addition to being used for BAO and $P_{1D}$ studies, FORESTFLOW has the potential to extend towards non-linear scales the current full-shape analyses of $\xi_{3D}$ (Cuceu et al. 2023; Gerardi et al. 2023) and $P_{3D}$ (Font-Ribera et al. 2018; de Belsunce et al. 2024; Horowitz et al. 2024), and can be used to interpret alternative three-dimensional statistics (Hui et al. 1999; Font-Ribera et al. 2018; Abdul Karim et al. 2024).

[1] Publicly available at https://github.com/igmhub/ForestFlow.

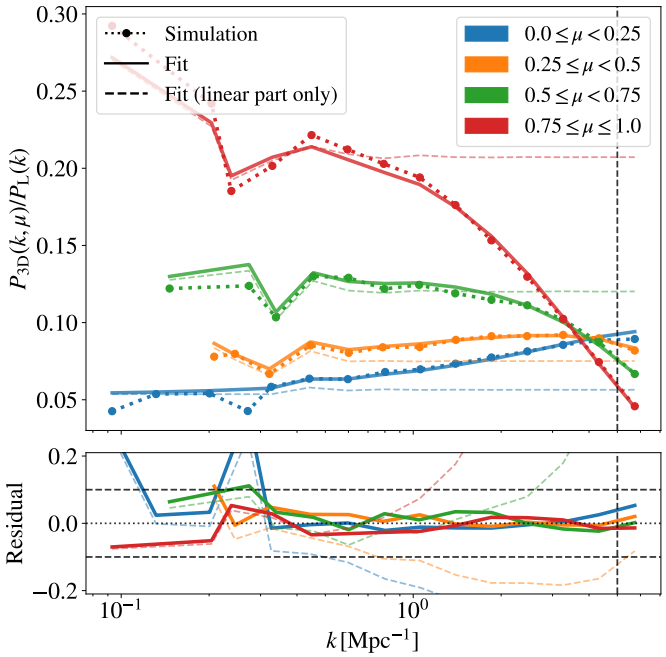


**Fig. 1.** Precision of the $P_{3D}$ model (see Eqs. 3 and 4) in reproducing measurements from the CENTRAL simulation at $z = 3$. In the top panel, dotted and solid lines show the ratio of simulation measurements and model predictions relative to the linear power spectrum, respectively. Dashed lines do so for the linear part of the best-fitting model ($D_{NL} = 1$). Line colors correspond to different $\mu$ wedges, and vertical dashed lines mark the minimum scale used for computing the best-fitting model, $k = 5\,\mathrm{Mpc}^{-1}$. The bottom panel displays the relative difference between the best-fitting model and simulation measurements. The overall precision of the model is 2% on scales in which simulation measurements are not strongly affected by cosmic variance ($k > 0.5\,\mathrm{Mpc}^{-1}$; see text).

The outline of this paper is as follows. We describe the emulation strategy in §2 and the suite of cosmological hydrodynamical simulations we use for training, how we measure $P_{3D}$ and $P_{1D}$ from these simulations, and our approach for computing the best-fitting parameters of the $P_{3D}$ analytic model to these statistics in §3. In §4 and 5, we present FORESTFLOW and evaluate its performance using multiple tests. In §6, we discuss some possible uses for this emulator, and we summarize our main results and conclude in §7.

Throughout this paper, all statistics and distances are in comoving units.

## 2. Emulation strategy

This paper aims to develop FORESTFLOW, a Lyman-$\alpha$ forest emulator predicting the 2 Lyman-$\alpha$ linear biases ($b_\delta$ and $b_\eta$) and 6 parameters capturing small-scale deviations of $P_{3D}$ from linear theory. We describe the emulation strategy in §2.1 and detail the input and output parameters of the emulator in §2.2.

### 2.1. $P_{3D}$ parametric model

We can express fluctuations in the Lyman-$\alpha$ forest flux as $\delta_F(\mathbf{s}) = \bar{F}^{-1}(\mathbf{s})F(\mathbf{s}) - 1$, where $F = \exp(-\tau)$ and $\bar{F}$ are the transmitted

flux fraction and its mean, respectively, $\tau$ is the optical depth to Lyman-$\alpha$ absorption, and $\mathbf{s}$ is the redshift-space coordinate. On linear scales, these fluctuations depend upon the matter field as follows (e.g.; McDonald 2003)

$$\delta_F = b_\delta\,\delta + b_\eta\,\eta, \tag{2}$$

where $\delta$ refers to matter density fluctuations, $\eta = -(a\,H)^{-1}(\partial v_\mathrm{r}/\partial r)$ stands for the dimensionless line-of-sight gradient of radial peculiar velocities, $a$ is the cosmological expansion factor, $H$ is the Hubble expansion factor, $v_\mathrm{r}$ is the radial velocity, and $r$ stands for the radial comoving coordinate. The linear bias coefficients $b_\delta$ and $b_\eta$ capture the response of $\delta_F$ to large-scale fluctuations in the $\delta$ and $\eta$ fields, respectively.

Following McDonald (2003), we decompose the three-dimensional power spectrum of $\delta_F$ into three terms

$$P_\mathrm{3D}(k,\,\mu) = (b_\delta + b_\eta\,f\,\mu^2)^2 D_\mathrm{NL}(k,\,\mu) P_\mathrm{lin}(k), \tag{3}$$

where $f = \mathrm{d}\log G/\mathrm{d}\log a$ is the logarithmic derivative of the growth factor $G$, $(b_\delta + b_\eta\,f\,\mu^2)^2$ accounts for linear biasing and large-scale redshift space distortions (Kaiser 1987; McDonald et al. 2000), $P_\mathrm{lin}$ is the linear matter power spectrum[2], and $D_\mathrm{NL}$ is a physically-motivated parametric correction accounting for the nonlinear growth of the density field, nonlinear peculiar velocities, thermal broadening, and pressure.

The large-scale behavior of $P_\mathrm{3D}$ is set by the bias coefficients $b_\delta$ and $b_\eta$ together with the linear power spectrum, and the latter can be computed using a Boltzmann solver (e.g.; Lewis et al. 2000; Lesgourgues 2011). Therefore, the emulation of the 2 Lyman-$\alpha$ linear biases enables predicting $P_\mathrm{3D}$ on arbitrarily large (linear) scales[3]. In contrast with direct emulation of power spectrum modes, this approach only requires simulations large enough for measuring the 2 Lyman-$\alpha$ linear biases precisely.

Making predictions for $P_\mathrm{3D}$ on small scales is more challenging than on large scales due to the variety of effects affecting this statistic on the nonlinear regime (e.g.; McDonald 2003). In this work, we describe small-scale effects using the physically-motivated Arinyo-i-Prats et al. (2015) parameterization

$$D_\mathrm{NL} = \exp\left\{\left(q_1\Delta^2 + q_2\Delta^4\right)\left[1 - \left(\frac{k}{k_\mathrm{v}}\right)^{a_\mathrm{v}}\mu^{b_\mathrm{v}}\right] - \left(\frac{k}{k_\mathrm{p}}\right)^2\right\}, \tag{4}$$

where $\Delta^2(k) \equiv (2\pi^2)^{-1}k^3 P_\mathrm{lin}(k)$ is the dimensionless linear matter power spectrum, $\mu$ is the cosine of the angle between the Fourier mode and the line of sight, and the free parameters $k_\mathrm{v}$ and $k_\mathrm{p}$ are in $\mathrm{Mpc}^{-1}$ units throughout this work. The terms involving $\{q_1,\,q_2\}$, $\{k_\mathrm{v},\,a_\mathrm{v},\,b_\mathrm{v}\}$, and $\{k_\mathrm{p}\}$ account for nonlinear growth, peculiar velocities and thermal broadening, and gas pressure, respectively. In (Givans et al. 2022), the authors used the previous expression to describe both $P_\mathrm{3D}$ and $P_\mathrm{1D}$ measurements down to highly nonlinear scales without the need for a shot-noise term. We will follow the same approach throughout this work; however, some authors advocate for such term (e.g.; Iršič & McQuinn 2018).

In the top panel of Fig. 1, dotted lines show the ratio of measurements from the CENTRAL simulation at $z = 3$ and the linear power spectrum, while solid lines do so for the best-fitting model to these measurements (Eqs. 3 and 4) and the linear power spectrum. See §3 for details about this simulation and the fitting procedure. The dashed lines depict the results for the best-fitting model when setting $D_\mathrm{NL} = 1$ after carrying out the fit; i.e., the behavior of the best-fitting model on linear scales. We can readily see that nonlinear growth isotropically increases the power with growing $k$, while peculiar velocities and thermal broadening suppress the power of parallel modes as $k$ increases. On even smaller scales, pressure takes over and causes an isotropic suppression. Nonlinear growth modifies the perpendicular power relative to linear theory by 10% for scales as large as $k = 0.5\,\mathrm{Mpc}^{-1}$, indicating that small-scale corrections are important for most of the scales sampled by our simulations. Nevertheless, our simulations are large enough to measure the 2 Lyman-$\alpha$ linear biases with percent precision (see Appendix A). Deviations from linear theory are less pronounced down to smaller scales for modes with $\mu \simeq 0.5$ because nonlinear growth and the combination of peculiar velocities and thermal broadening tend to cancel each other out.

On the largest scales, we find strong variations between consecutive $k$-bins for the same $\mu$-wedge. Some of these oscillations are driven by differences in the average value of $\mu$ between consecutive bins due to the limited number of modes entering each bin on large scales. To ensure an accurate comparison between simulation measurements and model predictions, we individually evaluate the $P_\mathrm{3D}$ model for all the modes within each $k-\mu$ bin from our simulation boxes. We then calculate the mean of the resulting distribution and assign this mean value to the bin, thereby mirroring the approach used to compute $P_\mathrm{3D}$ measurements from the simulations. This process is crucial for large scales where the number of modes is small and nonlinear scales where the dependence of the number of modes with $k$ is strong. We follow the same approach to evaluate the $P_\mathrm{3D}$ model throughout this work.

After accounting for the previous effect, the best-fitting model reproduces most large-scale oscillations. However, we can readily see a fluctuation at $k \simeq 0.25\,\mathrm{Mpc}^{-1}$ in the $0 < \mu < 0.25$ wedge that it is not captured by the model. The difference between model predictions and simulation measurements for the bins adjacent to this one is approximately zero, suggesting that this fluctuation is caused by cosmic variance. We characterize the impact of this source of uncertainty in Appendix A, concluding that it can induce up to 10% errors on scales larger than $k = 0.5\,\mathrm{Mpc}^{-1}$. Consequently, cosmic variance hinders our ability to evaluate the model's performance on the largest scales shown. However, this does not necessarily indicate a decrease in model precision on scales larger than $k = 0.5\,\mathrm{Mpc}^{-1}$; rather, our simulations are simply not large enough to accurately assess the model's precision on such scales. In the bottom panel of Fig. 1, we can see that the average precision of the model is 2% for $k > 0.5\,\mathrm{Mpc}^{-1}$, supporting the use of Eq. 4 for capturing small-scale deviations from linear theory.

### 2.2. Input and output parameters

In addition to the density and velocity fields, the Lyman-$\alpha$ forest depends upon the ionization and thermal state of the IGM (e.g.; McDonald 2003). Following Pedersen et al. (2021), we use 6 parameters to describe the dependency of this observable with cosmology and IGM physics:

- Amplitude ($\Delta^2_\mathrm{p}$) and slope ($n_\mathrm{p}$) of the linear matter power spectrum on small scales. These are defined as

$$\Delta^2_\mathrm{p}(z) = (2\pi^2)^{-1}k^3 P_\mathrm{lin}(k_\mathrm{p}, z), \tag{5}$$

$$n_\mathrm{p}(z) = (\mathrm{d}\log P_\mathrm{lin}/\mathrm{d}\log k)\,|_{k=k_\mathrm{p}}, \tag{6}$$

[2] This is the linear power spectrum of cold dark matter and baryons even for cosmologies with massive neutrinos.

[3] Aside from nonlinear effects affecting the position and damping of BAO.

where we use $k_p = 0.7\,\mathrm{Mpc}^{-1}$ as the pivot scale because it is at the center of the range of interest for DESI small-scale studies. These parameters capture multiple physical effects modifying the linear power spectrum on small scales (see Pedersen et al. 2021, for a detailed discussion), including cosmological parameters such as the amplitude ($A_s$) and slope ($n_s$) of the primordial power spectrum, the Hubble parameter, and the matter density ($\Omega_M$), or ΛCDM extensions such as curvature and massive neutrinos. The advantage of using this parameterization rather than ΛCDM parameters is twofold. First, we reduce the dimensionality of the emulator input, which decreases the number of simulations required for precise training. Second, the resulting emulator has the potential for making precise predictions for variations in cosmological parameters and ΛCDM extensions not considered in the training set (Pedersen et al. 2021; Pedersen et al. 2023; Cabayol-Garcia et al. 2023). Note that we do not consider cosmological parameters capturing changes in the growth rate or expansion history because the Lyman-$\alpha$ forest probes cosmic times during which the universe is practically Einstein de-Sitter, and both vary very little with cosmology in this regime.

- Mean transmitted flux fraction ($\bar{F}$). It depends on the intensity of the cosmic ionizing background and evolves strongly with redshift. One of the advantages of using this parameter is that it encodes the majority of the redshift dependence of the signal, serving as a proxy for cosmic time.
- Amplitude and slope of the temperature-density relation. The thermal state of the IGM can be approximated by a power law on the densities probed by the Lyman-$\alpha$ forest (Lukić et al. 2015): $T_0\Delta_b^{\gamma-1}$, where $\Delta_b$ is the baryon overdensity, $T_0$ is the gas temperature at mean density, and $\gamma - 1$ is the slope of the relation. These parameters influence the ionization of the IGM, which is captured by $\bar{F}$, and the thermal motion of gas particles, which causes Doppler broadening that suppresses the parallel power. Instead of using $T_0$ as an emulator parameter, we follow Pedersen et al. (2021) and use the thermal broadening scale in comoving units. First, we express the thermal broadening in velocity units as $\tilde{\sigma}_T = 9.1(T_0[\mathrm{K}]/10^4)^{1/2}$, and then we convert it to comoving units, $\sigma_T = \tilde{\sigma}_T(1+z)H^{-1}$.
- Pressure smoothing scale. Gas pressure supports baryons on small scales, leading to a strong isotropic power suppression in this regime. The characteristic smoothing scale depends upon the entire thermal history of the gas (Gnedin & Hui 1998), and we parameterize its effect using the pressure smoothing scale in comoving units of $\mathrm{Mpc}^{-1}$, $k_F$ (see Pedersen et al. (2021) for more details).

Our emulator predicts the 8 free parameters of the $P_{3D}$ model introduced by Eqs. 3 and 4, $\mathbf{y} = \{b_\delta, b_\eta, q_1, q_2, k_v, a_v, b_v, k_p\}$, as a function of the previous 6 parameters, $\mathbf{x} = \{\Delta_p^2, n_p, \bar{F}, \sigma_T, \gamma, k_F\}$. In the three next sections, we generate the training data of the emulator, discuss its implementation, and evaluate its precision.

## 3. Training and testing set

In this section, we describe how we generate the training and testing data of our emulator. In §3.1, we present a suite of cosmological hydrodynamical simulations from which we generate mock Lyman-$\alpha$ forest measurements, and we detail our approach for extracting $P_{3D}$ and $P_{1D}$ measurements from these simulations in §3.2. In §3.3, we compute the best-fitting parameters of the model introduced by Eqs. 3 and 4 to measurements of these statistics, and we evaluate the performance of the fits in §3.4.

### 3.1. Simulations

We extract Lyman-$\alpha$ forest measurements from a suite of simulations run with MP-GADGET[4] (Feng et al. 2018; Bird et al. 2019), a massively scalable version of the cosmological structure formation code GADGET-3 (last described in Springel 2005). This suite of simulations was first presented and used in Pedersen et al. (2021); we briefly describe it next. Each simulation tracks the evolution of $768^3$ dark matter and baryon particles from $z = 99$ to $z = 2$ inside a box of $L = 67.5$ Mpc on a side, producing as output 11 snapshots uniformly spaced in redshift between $z = 4.5$ and 2. This configuration ensures convergence for $P_{1D}$ measurements down to $k_\parallel = 4\,\mathrm{Mpc}^{-1}$ (the smallest scale used in this work) at $z = 2$ and less than 10% errors for this scale at $z = 4$. For more details, see the box size and mass resolution tests carried out in Bolton et al. (2017).

Two realizations were run for each combination of cosmological and astrophysical parameters using the "fixed-and-paired" technique (Angulo & Pontzen 2016; Pontzen et al. 2016), which significantly reduces cosmic variance for multiple observables, including the Lyman-$\alpha$ forest (Villaescusa-Navarro et al. 2018; Anderson et al. 2019). The initial conditions were generated using the following configuration of MP-GENIC (Bird et al. 2020): initial displacements produced using the Zel'dovich approximation and baryons and dark matter initialized on an offset grid using species-specific transfer functions. Some studies have suggested that this configuration might lead to incorrect evolution of linear modes (Bird et al. 2020). However, in a recent study, Khan et al. (2024) showed that variations in the specific settings of MP-GENIC initial conditions have a minimal impact on $P_{1D}$ measurements across the range of redshifts and scales used in this work.

To increase computational efficiency, the simulations utilize a simplified prescription for star formation that turns regions of baryon overdensity $\Delta_b > 1000$ and temperature $T < 10^5$ K into collisionless stars (e.g.; Viel et al. 2004), implement a spatially uniform ultraviolet background (Haardt & Madau 2012), and do not consider active galactic nuclei (AGN) feedback (e.g.; Chabanier et al. 2020). These approximations are justified because we focus on emulating the Lyman-$\alpha$ forest in the absence of astrophysical contaminants like AGN feedback, damped Lyman-alpha absorbers (DLAs), or metal absorbers, and we will model these before comparing our predictions with observational measurements (e.g.; McDonald et al. 2005; Palanque-Delabrouille et al. 2015, 2020).

We train FORESTFLOW using data from 30 fixed-and-paired simulations spanning combinations of cosmological and astrophysical parameters selected according to a Latin hypercube (McKay et al. 1979); we refer to these as TRAINING simulations hereafter. The Latin hypercube spans the parameters $\{\Delta_p^2(z = 3), n_p(z = 3), z_H, H_A, H_S\}$, where we use $z = 3$ because it is approximately at the center of the range of interest for DESI studies (Ravoux et al. 2023; Karaçaylı et al. 2024), $z_H$ is the midpoint of hydrogen reionization, and the last two parameters rescale the He II photoheating rate $\epsilon_0$ as $\epsilon = H_A\Delta_b^{H_S}\epsilon_0$ (Oñorbe et al. 2017). Cosmological parameters were generated within the ranges $\Delta_p^2(z = 3) \in [0.25, 0.45]$, $n_p(z = 3) \in [-2.35, -2.25]$ by exploring values of the amplitude and slope of the primordial power spectrum within the intervals $A_s \in [1.35, 2.71] \times 10^{-9}$ and $n_s \in [0.92, 1.02]$. Any other ΛCDM parameter was held

[4] https://github.com/MP-Gadget/MP-Gadget/

fixed to approximately Planck Collaboration et al. (2020) values: dimensionless Hubble parameter $h = 0.67$, physical cold dark matter density $\Omega_c h^2 = 0.12$, and physical baryon density $\Omega_b h^2 = 0.022$. As for the IGM parameters, these explored the ranges $z_H \in [5.5, 15]$, $H_A \in [0.5, 1.5]$, and $H_S \in [0.5, 1.5]$. All simulation pairs use the same set of initial Fourier phases, and thus cosmic variance affects in approximately the same way all combinations of input parameters.

We evaluate different aspects of the emulation strategy using 6 fixed-and-paired simulations with cosmological and astrophysical parameters not considered in the TRAINING simulations:

- The CENTRAL simulation uses cosmological and astrophysical parameters at the center of the TRAINING parameter space: $A_s = 2.01 \times 10^{-9}$, $n_s = 0.97$, $z_H = 10.5$, $H_A = 1$, and $H_S = 1$. We use this simulation for an out-of-sample test of the emulator's performance under optimal conditions, as the precision of machine-learning models typically decreases as we move closer to the border of the convex hull set by the training set.
- The SEED simulation uses the same parameters as the CENTRAL simulation while considering a different distribution of initial Fourier phases. Given that all TRAINING simulations use the same initial Fourier phases, SEED is useful to evaluate the impact of cosmic variance in the training set on FORESTFLOW predictions.
- The GROWTH, NEUTRINOS, and CURVED simulations adopt the same values of $\Delta^2_p(z = 3)$, $n_p(z = 3)$, physical cold dark matter and baryonic densities, and astrophysical parameters as the CENTRAL simulation. However, the GROWTH simulation uses 10% larger Hubble parameter ($h = 0.74$) and 18% smaller matter density ($\Omega_M = 0.259$) while using the same value of $\Omega_M h^2$ as the TRAINING simulations, the NEUTRINOS simulation includes massive neutrinos ($\sum m_\nu = 0.3$ eV), and the CURVED simulation considers an open universe ($\Omega_k = 0.03$). The NEUTRINOS and CURVED simulations also modify the value of the cosmological constant while holding fixed $h$ to compensate for the increase in the matter density and the addition of curvature, respectively. The testing simulations address the precision of the emulation strategy for cosmologies not included in the training set.
- The REIONISATION simulation uses the same cosmological parameters as the CENTRAL simulation while implementing a distinct helium ionization history relative to the CENTRAL and TRAINING simulations (Puchwein et al. 2019). The main difference between the ionization histories of these simulations is that the one implemented in the REIONISATION simulation peaks at a later time than the others, leading to a significantly different thermal history. The REIONISATION simulation therefore tests the emulator's performance for thermal histories not considered in the TRAINING simulations.

### 3.2. Simulating Lyman-$\alpha$ forest data

To extract Lyman-$\alpha$ forest measurements from each simulation, we first select one of the simulation axes as the line of sight and displace the simulation particles from real to redshift space along this axis. Then, we compute the transmitted flux fraction along $768^2$ uniformly-distributed line of sights along this axis using FSFE[5] (Bird 2017); these lines of sight are commonly known as skewers. The resolution of the skewers is set to 0.05 Mpc, which is enough to resolve the thermal broadening and pressure scales, and are spaced by 0.09 Mpc in the transverse direction.

[5] https://github.com/sbird/fake_spectra.

We checked that $P_{3D}$ and $P_{1D}$ measurements within the range of interest (see §3.3) do not vary by increasing the line-of-sight resolution or the transverse sampling. After that, we repeat the previous steps for the three simulation axes to extract further cosmological information, as each simulation axis samples the velocity field in a different direction. Finally, we scale the effective optical depth of the skewers to 0.90, 0.95, 1.05, and 1.10 times its original value (see Lukić et al. 2015, for more details about this approach), which is equivalent to running simulations with different values of the UV background.

Using this data as input, we measure $P_{3D}$ by first computing the three-dimensional Fourier transform of the skewers. Then, we take the average of the square norm of all modes within 20 logarithmically-spaced bins in wavenumber $k$ from the fundamental mode of the box, $k_{min} = 2\pi L^{-1} \simeq 0.09\,\mathrm{Mpc}^{-1}$, to $k_{max} = 40\,\mathrm{Mpc}^{-1}$ and 16 linearly-spaced bins in the cosine of the angle between Fourier modes and the line of sight from $\mu = 0$ to 1. We measure $P_{1D}$ by first computing the one-dimensional Fourier transform of each skewer without applying any binning, and then by taking the average of the square norm of all these Fourier transforms.

We carry out these measurements for the 30 TRAINING and 6 test fix-and-paired simulations, ending up with 2 (opposite Fourier phases) × 3 (simulation axes) × 11 (snapshots) × 5 (mean flux rescalings) = 330 measurements per simulation. To reduce cosmic variance, we compute the average of measurements from different axes and phases of fixed-and-paired simulations, decreasing the number of measurements per simulation to 55. The training and testing sets of FORESTFLOW are thus comprised of 1650 and 330 Lyman-$\alpha$ power spectrum measurements, respectively. All these measurements are publicly available at https://github.com/igmhub/LaCE.

### 3.3. Fitting the parametric model

To generate training and testing data for our emulator, we compute the best-fitting parameters of Eqs. 3 and 4 to measurements from the simulations described in §3.1. We fit the model using $P_{3D}$ measurements from $k = 0.09$ to $5\,\mathrm{Mpc}^{-1}$ and $P_{1D}$ measurements from $k_\parallel = 0.09$ to $4\,\mathrm{Mpc}^{-1}$. The size of our simulation boxes determines the largest scales used, while the maximum wavenumbers are set by the smallest scales measured by (Ravoux et al. 2023; Karaçaylı et al. 2024). We remind the reader that the large-scale behavior of $P_{3D}$ is set by the 2 Lyman-$\alpha$ linear biases (see Eq. 3); consequently, the model can make accurate predictions for $P_{3D}$ on arbitrarily large (linear) scales as long as these 2 parameters are measured precisely.

We compute the best-fitting value of model parameters $\mathbf{y} = \{b_\delta, b_\eta, q_1, q_2, k_v, a_v, b_v, k_p\}$ to simulation measurements by minimizing the pseudo-$\chi^2$:

$$\chi^2(\mathbf{y}) = \sum_i^{M_{3D}} w_{3D} \left[P_{3D}^{data}(k_i, \mu_i) - P_{3D}^{model}(k_i, \mu_i, \mathbf{y})\right]^2 + \sum_i^{M_{1D}} w_{1D} \left[P_{1D}^{data}(k_{\parallel,i}) - P_{1D}^{model}(k_{\parallel,i}, \mathbf{y})\right]^2, \quad (7)$$

where $M_{3D} = 164$ and $M_{1D} = 42$ are the number of $P_{3D}$ and $P_{1D}$ bins employed in the fit, respectively, the superscripts data and model refer to simulation measurements and model predictions, and $w_{3D}$ and $w_{1D}$ weigh the fit. We use the Nelder-Mead algorithm implemented in the routine MINIMIZE of SCIPY (Virtanen et al. 2020) to carry out the min-

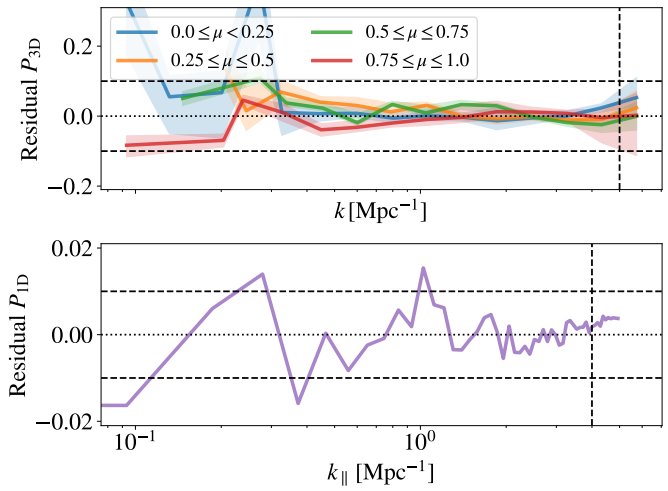


**Fig. 2.** Precision of the parametric model (see Eqs. 3 and 4) in reproducing $P_{3D}$ and $P_{1D}$ measurements from all the TRAINING simulations. Lines and shaded areas show the mean and standard deviation of the relative difference between simulation measurements from the 1650 snapshots of the TRAINING simulations and best-fitting models to these, respectively. The precision of the model in recovering $P_{3D}$ and $P_{1D}$ is 2.4 and 0.6%, respectively, on scales not strongly affected by cosmic variance.

imization [6]. The results of the fits are publicly accessible at https://github.com/igmhub/ForestFlow.

Ideally, we would use the covariance of $P_{3D}$ and $P_{1D}$ measurements as weight in the previous expression; however, we lack multiple realizations of the same simulation with different distributions of initial Fourier phases needed to estimate this covariance. In addition, its theoretical estimation is not straightforward (Maion et al. 2022). Instead, we disregard correlations between $P_{3D}$ and $P_{1D}$ and weigh these by $w_{3D} = N_{3D}(k,\mu)/(1+\mu^2)^2$ and $w_{1D} = \alpha(1+k_\parallel/k_0)^2$, where $N_{3D}$ is the number of modes in each $k-\mu$ bin and $k_0 = 2\,\mathrm{Mpc}^{-1}$. The terms involving $N_{3D}$, $\mu$, and $k_0$ attempt to ensure an unbiased fit of $P_{3D}$ and $P_{1D}$ across the full range of scales used. The parameter $\alpha = 8000$ controls the relative weight of $P_{3D}$ and $P_{1D}$ in the fit, and its value is motivated by the different impact of cosmic variance on these (see Appendix A).

We expect significant correlations between the best-fitting value of the parameters to measurements from relatively small simulation boxes. As shown by Arinyo-i-Prats et al. (2015), these correlations are especially significant for the parameters accounting for nonlinear growth of structure, $q_1$ and $q_2$. Givans et al. (2022) advocated for setting $q_2 = 0$ since this parameter is not necessary for describing $P_{3D}$ at $z = 2.8$. However, we find non-zero values of this parameter indispensable for describing $P_{3D}$ at redshifts below $z = 2.5$. This is not surprising since the gravitational evolution of density perturbations becomes increasingly more nonlinear as cosmic time progresses.

### 3.4. Precision of the model

In the previous section, we compute the best-fitting parameters of the $P_{3D}$ model to measurements from the TRAINING simulations. Two main sources of uncertainty can affect these fits: model inaccuracies and cosmic variance. The first relates to using a model with a limited number of free parameters to describe Lyman-$\alpha$ clustering, while the second comes from the limited size of the TRAINING simulations. In our case, cosmic variance is amplified because all the TRAINING simulations use the same initial distribution of Fourier phases, meaning all simulations are subject to the same large-scale noise. We study this source of uncertainty in Appendix A, where we compare the best-fitting models to two simulations whose only difference is in their initial distribution of Fourier phases. We proceed to study model inaccuracies next.

In Fig. 2, we evaluate the performance of the parametric model in reproducing $P_{3D}$ and $P_{1D}$ measurements from the 1650 snapshots of the TRAINING simulations. As discussed in §2.1, cosmic variance limits our ability to evaluate the precision of the model for $P_{3D}$ on scales larger than $k = 0.5\,\mathrm{Mpc}^{-1}$; therefore, we quote the model precision from this scale down to the smallest scale used in the fit, $k = 5\,\mathrm{Mpc}^{-1}$. In contrast, since cosmic variance has a much smaller impact on $P_{1D}$, we evaluate the model performance for this statistic using all scales considered in the fit, $0.09 < k_\parallel[\,\mathrm{Mpc}^{-1}] < 4$. We adopt the same approach when evaluating the precision of the emulator in §5. Under these considerations, the overall precision of the model is 2.4 and 0.6% for $P_{3D}$ and $P_{1D}$, respectively.

Given that we estimate the precision of the model using our simulations, the previous numbers account for both the limited flexibility of the $P_{3D}$ model and cosmic variance. As discussed in Appendix A, the impact of cosmic variance on measurements of $P_{3D}$ and $P_{1D}$ from our simulations is 1.3 and 0.5%, respectively. This indicates that the limited flexibility of the $P_{3D}$ model introduces additional errors of 1.1 and 0.1% on these statistics beyond cosmic variance. We can thus conclude that the $P_{3D}$ model accurately reproduces $P_{3D}$ and $P_{1D}$ over the full range of scales used in the fit, $0.09 < k[\,\mathrm{Mpc}^{-1}] < 5$ and $0.09 < k_\parallel[\,\mathrm{Mpc}^{-1}] < 4$.

## 4. ForestFlow

In this section we present FORESTFLOW, an emulator based on conditional normalizing flows that predicts the parameters of the $P_{3D}$ model introduced by Eqs. 3 and 4 as a function of parameters capturing the dependence of the Lyman-$\alpha$ forest on cosmology and IGM physics. We detail its architecture and implementation in §4.1 and 4.2, respectively.

### 4.1. Conditional normalizing flows

Normalizing flows (NFs; Jimenez Rezende & Mohamed 2015) are a class of machine-learning generative models designed to predict complex distributions by applying a sequence of bijective mappings to simple base distributions. A natural extension to this framework is conditional NFs (cNFs; Winkler et al. 2019; Papamakarios et al. 2019), a type of NFs that condition the mapping between the base and target distributions on a series of input variables. Given an input $\mathbf{x} \in X$ and target $\mathbf{y} \in Y$, cNFs predict the conditional distribution $p_{Y|X}(\mathbf{y}|\mathbf{x})$ by applying a parametric, bijective mapping $f_\phi : Y \times X \to Z$ to a base distribution $p_Z(\mathbf{z})$ as follows

$$p_{Y|X}(\mathbf{y}|\mathbf{x}) = p_Z(f_\phi(\mathbf{y},\mathbf{x}))\left|\frac{\partial f_\phi(\mathbf{y},\mathbf{x})}{\partial \mathbf{y}}\right|, \quad (8)$$

where $\phi$ are the parameters of the mapping, while the last term of the previous equation is the Jacobian determinant of the mapping. In FORESTFLOW, the input is given by the parameters capturing the dependence of the Lyman-$\alpha$ forest on cosmology and IGM

[6] To ensure that this routine does not get stuck in a local minimum, we checked that the likelihood is unimodal in all cases using the Affine Invariant Markov chain Monte Carlo Ensemble sampler EMCEE (Foreman-Mackey et al. 2013).

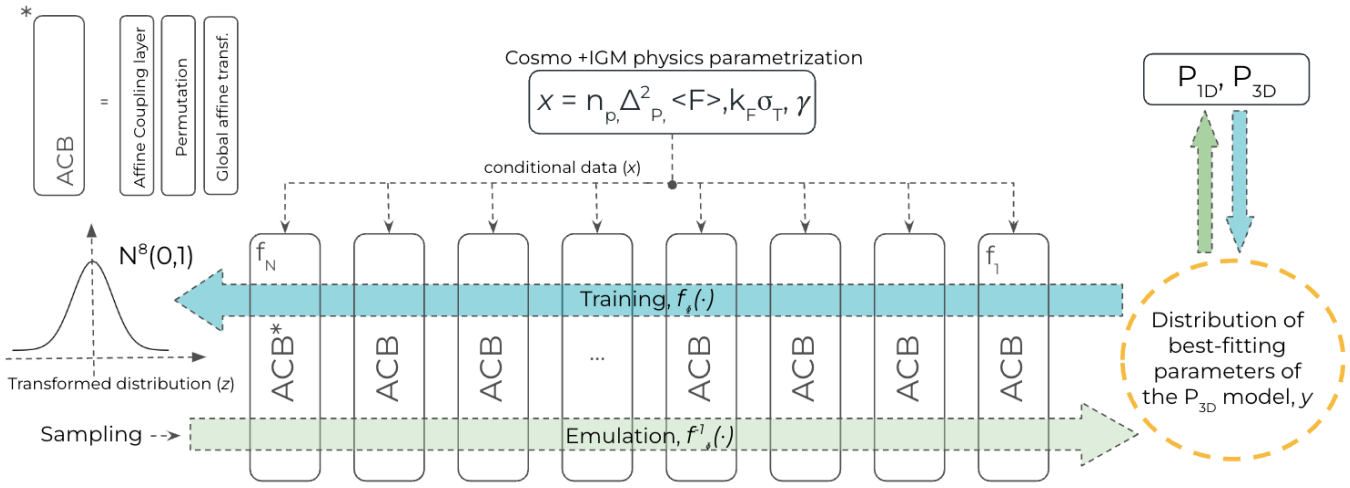


**Fig. 3.** Architecture of FORESTFLOW, a Lyman-$\alpha$ forest emulator based on conditional Normalizing Flows. The blue arrow indicates the training direction, where FORESTFLOW optimizes a bijective mapping between the best-fitting parameters of the $P_{3D}$ model to measurements from the TRAINING simulations and an 8-dimensional Normal distribution. The mapping is conditioned on cosmology and IGM physics and performed using 12 consecutive affine coupling blocks. The green arrow denotes the emulation direction, where the emulator applies the inverse of the mapping to random samples from the base distribution to predict the value of the $P_{3D}$ model parameters. Outside the cNF, FORESTFLOW introduces these parameters in Eq. 3 and 1 to obtain predictions for $P_{3D}$ and $P_{1D}$, respectively.

physics, $\mathbf{x} = \{\Delta^2_p,\, n_p,\, \bar{F},\, \sigma_T,\, \gamma,\, k_F\}$, the target by the parameters of the $P_{3D}$ model, $\mathbf{y} = \{b_\delta,\, b_\eta,\, q_1,\, q_2,\, k_v,\, a_v,\, b_v,\, k_p\}$, and the base distribution is an 8-dimensional Normal distribution $N^8(0,1)$, where the dimension is determined by the number of $P_{3D}$ model parameters.

Once trained, cNFs are a generative process from $\mathbf{x}$ to $\mathbf{y}$. In our implementation, FORESTFLOW first samples randomly from the base distribution, and then it passes this realization through a sequence of mappings conditioned on a particular combination of cosmology and IGM parameters, $\tilde{\mathbf{y}} = f_\phi^{-1}(p_Z(\mathbf{z}), \mathbf{x})$, ending up with a prediction for the value of the $P_{3D}$ parameters. Repeating this process multiple times, the FORESTFLOW yields a distribution of $P_{3D}$ parameters $p_{\tilde{Y}|X}$ that, for a sufficiently large number of samples, approaches the target distribution $p_{Y|X}$. The breadth of this distribution captures uncertainties arising from the limited size of the training set. Finally, outside the cNF, we use each combination of $P_{3D}$ parameters to evaluate Eqs. 3 and 1, obtaining predictions and uncertainties for $P_{3D}$ and $P_{1D}$.

The main challenge when using cNFs is finding the mapping between the target and the base distribution, typically done using an $N$-layer neural network with bijective layers. This process runs in reverse relative to the generating process: we start by applying the mapping $f_\phi$ to the target data $\mathbf{y}$ conditioned on the input $\mathbf{x}$, yielding $\mathbf{z}$. Then, we optimize the model parameters by minimizing the loss function

$$\mathcal{L} = \frac{1}{2}\sum \mathbf{z}^2 - \log\left|\frac{\delta f_\phi(\mathbf{y},\mathbf{x})}{\delta \mathbf{y}}\right|. \tag{9}$$

We carry out this optimization process using stochastic gradient descent applied to minibatches, a methodology commonly employed for training neural networks.

### 4.2. Implementation

Neural Autoregressive Flows (Huang et al. 2018) use a series of invertible univariate operations to build a bijective transformation between a conditional distribution and a base distribution. In FORESTFLOW, we create a bijective mapping between the best-fitting parameters of the $P_{3D}$ model and an 8-dimensional Normal distribution by applying $N_{ACB} = 12$ consecutive Affine-Coupling Block (ACB; Dinh et al. 2016) conditioned on cosmology and IGM physics. The transformation goes from the best-fitting parameters of the $P_{3D}$ model to the base distribution when training the model, and in the opposite direction when evaluating it.

Each ACB conducts a series of operations $g_{i,\tilde{\phi}_i}$ on its input data $\mathbf{w}_i$, with $i$ going from 1 to $N_{ACB}$ and $\tilde{\phi}_i$ standing for the parameters of the transformation. First, it splits the input data into two subsamples with approximately the same number of elements, $\mathbf{w'}_i$ and $\mathbf{w''}_i$. Then, it applies an affine transformation to the first subsample $\mathbf{w'}_i$

$$T(\mathbf{w'}_i) = \alpha_i\, \mathbf{w'}_i + \beta_i, \tag{10}$$

where $\alpha_i$ and $\beta_i$ are neural networks with a single hidden layer of 128 neuron units. Third, the ACB merges the output from the affine transformation and the unchanged subsample, and then it applies a permutation layer to randomly rearrange these elements, obtaining $\tilde{\mathbf{w}}_i$. Fourth, the ACB applies an affine transformation to this sample, $\tilde{T}(\tilde{\mathbf{w}}_i)$. The first and second affine transformations involve a subset of the training set and the entire training set, respectively, enabling the model to capture local and global features.

In Fig. 3, we show the architecture of FORESTFLOW. The blue arrow indicates the training direction, while the green arrow depicts the emulation direction. In the training direction, the input to the first ACB, $\mathbf{u}_1 = \mathbf{w}_1$, is a 1650-dimensional array composed of 14-dimensional vectors, where 1650 is the number of simulation snapshots in the training set. Each vector includes the 8 best-fitting $P_{3D}$ model parameters to each snapshot and the 6 parameters describing the cosmology and IGM physics of this snapshot. The input to the $i$ ACB, $\mathbf{u}_i$, is a 1650-array containing 14-dimensional vectors with the output of the $i-1$ ACB and, once again, the 6 parameters describing the cosmology and IGM physics of each snapshot. Each ACB applies a transformation $f_{i,\phi_i} = g_{i,\tilde{\phi}_i}$, and the consecutive application of all ACBs results

in the mapping between the target and the base distributions $\mathbf{z} = f_\phi(\mathbf{y}, \mathbf{x})$, where $f_\phi = \prod_{i=1}^{N_{\rm ACB}} f_{i,\phi_i}$.

In the emulation direction, the input to the first ACB, $\mathbf{v}_1 = \mathbf{w}_1$, is a 14-dimensional vector containing random draws from an 8-dimensional Normal distribution and the 6 parameters describing the cosmology and IGM physics for which we want to obtain predictions. As in the training direction, the input to each subsequent ACB relies on the output from the previous ACB, each conditioned on cosmology and IGM physics. The ACBs apply the transformations $f^{-1}_{i,\phi_i} = g_{i,\tilde{\phi}_i}$, which are the inverse of the corresponding transformations in the training direction, $f_{i,\phi_i}$. FORESTFLOW makes predictions for $P_{\rm 3D}$ model parameters by applying the composition of the inverse of all ACBs to random samples from the base distribution, $\tilde{\mathbf{y}} = f^{-1}_\phi(p_Z(\mathbf{z}), \mathbf{x})$, where $f^{-1}_\phi = \prod_{i=1}^{N_{\rm ACB}} f^{-1}_{i,\phi_i}$.

We implement the emulator within the `FreIA` framework (Ardizzone et al. 2018-2022), which uses `PyTorch` (Ansel et al. 2024) in the backend. FORESTFLOW is trained by minimizing Eq. 9 using an `Adam` optimizer (Kingma & Ba 2014) for 300 epochs with an initial learning rate of $10^{-3}$. We use the `Optuna` framework (Akiba et al. 2019) to select the number of ACBs and epochs, as well as the value of the learning rate. First, `Optuna` trains FORESTFLOW for a particular combination of these hyperparameters. Then, it computes the average value of Eq. 7 for all simulations in the training set. After that, depending on the goodness of the fit to $P_{\rm 3D}$ and $P_{\rm 1D}$ measurements, `Optuna` selects a new value of the hyperparameters. We iterate with `Optuna` 50 times through a hyperparameter grid, selecting the hyperparameters that yield the highest precision. We checked that the precision of the emulator depends weakly on small variations in the value of the hyperparameters. FORESTFLOW is publicly available at https://github.com/igmhub/ForestFlow.

## 5. Emulator performance

In this section, we evaluate the performance of FORESTFLOW. In §5.1, we assess its performance throughout the parameter space of the training set. Then, in §5.2, we examine the precision of the emulator using simulations with cosmologies and IGM models not included in the training set.

### 5.1. Throughout the parameter space of the training set

In this section, we evaluate the precision of FORESTFLOW in recovering the 2 Lyman-$\alpha$ linear biases, which determine the behavior of $P_{\rm 3D}$ on linear scales, as well as $P_{\rm 3D}$ and $P_{\rm 1D}$ measurements from simulations on the intervals $0.5 < k\,[\rm Mpc^{-1}] < 5$ and $0.09 < k_\parallel\,[\rm Mpc^{-1}] < 4$, respectively. These are the ranges of scales used when fitting the parametric model in §3 that are not strongly affected by cosmic variance (see §2.1). We begin by assessing the precision at the center of the training set, where machine-learning methods typically perform best, and then extend our evaluation across the entire input parameter space.

In Fig. 4, we compare measurements of $P_{\rm 3D}$ and $P_{\rm 1D}$ from the CENTRAL simulation at $z = 3$ with FORESTFLOW predictions. Dotted lines show simulation measurements, while solid lines and shaded areas display the average and 68% credible interval of FORESTFLOW predictions, respectively. We characterize the accuracy of the credible intervals in Appendix B. As we can see, the emulator captures the amplitude and scale-dependence of $P_{\rm 3D}$ and $P_{\rm 1D}$ precisely. To better characterize the emulator's performance, we compute the average precision of FORESTFLOW in recovering measurements from CENTRAL across redshift. We find that it is 1.2 and 0.3% for $b_\delta$ and $b_\eta$, respectively, which translates into 1.1 and 1.2% for perpendicular and parallel $P_{\rm 3D}$ modes on linear scales, and 2.6 and 0.8% for $P_{\rm 3D}$ and $P_{\rm 1D}$. Note that cosmic variance hinders our ability to test the precision of the model; however, this does not necessarily indicate a decrease in model precision for $P_{\rm 3D}$ on scales larger than $k = 0.5\,\rm Mpc^{-1}$.

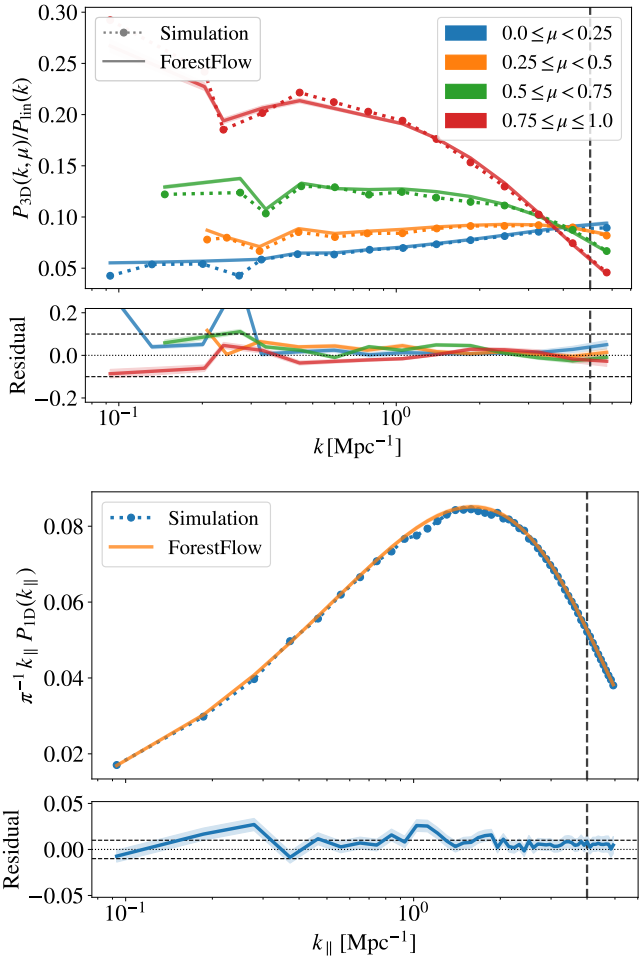


**Fig. 4.** Precision of the emulator in recovering $P_{\rm 3D}$ and $P_{\rm 1D}$ measurements from the CENTRAL simulation at $z = 3$. Dotted lines show measurements from simulations, solid lines and shaded areas display the average and 68% credible interval of FORESTFLOW predictions, respectively, and vertical dashed lines indicate the minimum scales considered for computing the training data of the emulator. The overall precision of the emulator in recovering $P_{\rm 3D}$ is 2.0% on scales not strongly affected by cosmic variance and 0.6% for $P_{\rm 1D}$.

We expect the emulator's efficiency to decrease away from the center of the input space. Ideally, we would have multiple test simulations covering the entire input space to evaluate the performance, but such simulations are unavailable. Instead, we conduct leave-one-out tests, which are widely used to assess the precision of an emulator when the number of training points is insufficient for out-of-sample tests (e.g.; Hastie et al. 2001). In a leave-one-out test, we optimize FORESTFLOW after removing a subsample from the training set; for example, all measurements from one of the TRAINING simulations. We then check the precision of the new emulator using the subsample held back. By repeating this process for other subsamples, we can estimate the performance of FORESTFLOW across the parameter space. Since each emulator is trained without using the entire dataset, leave-one-out tests provide a lower bound on emulator performance. Additionally,

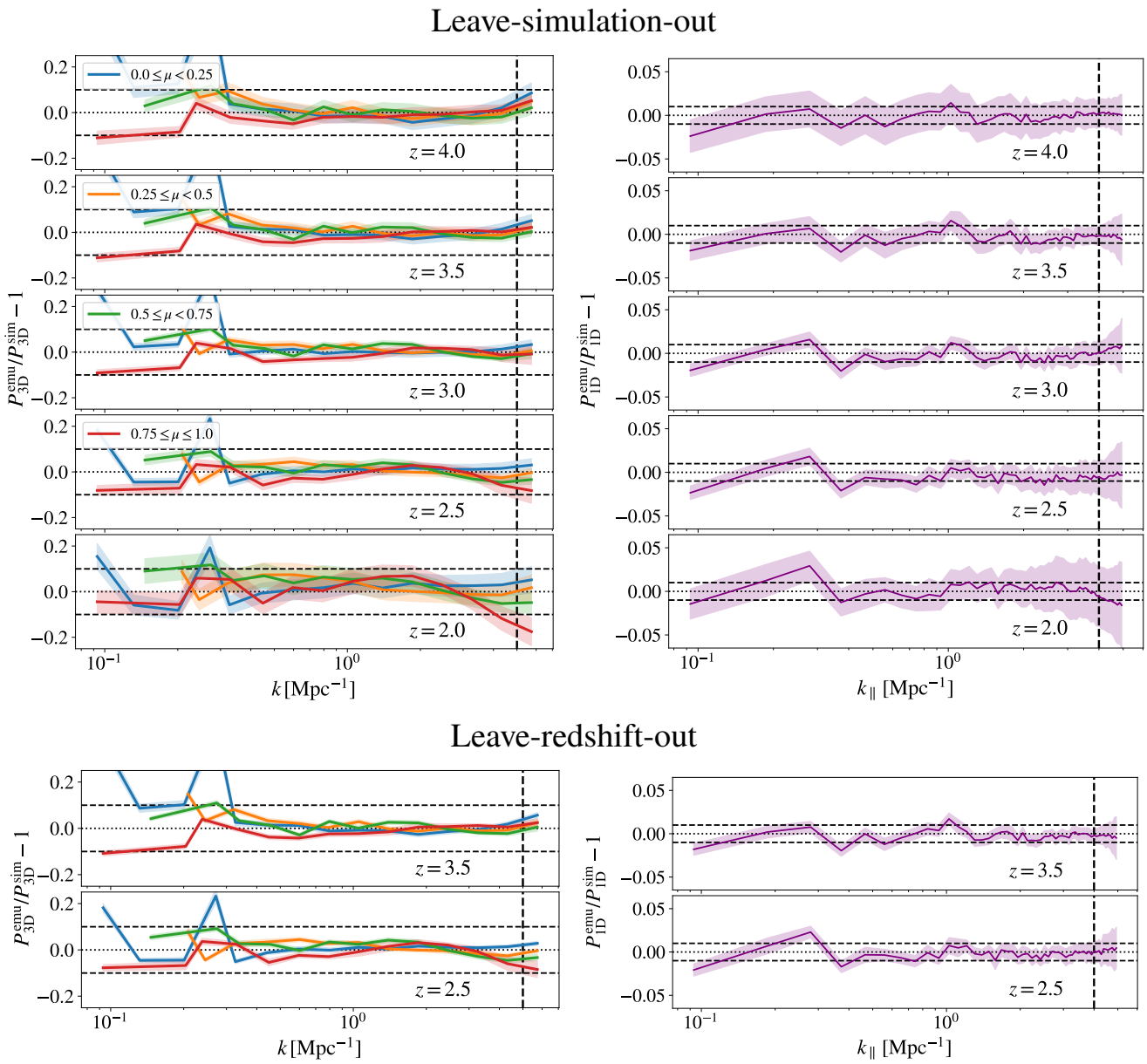


**Fig. 5.** Precision of the emulator across the input parameter space estimated via leave-simulation-out (top panels) and leave-redshift-out tests (bottom panels). **Top panels.** Each leave-simulation-out test involves training 1 independent emulator with measurements from 29 distinct simulations, and then using the measurements from the remaining simulation as the validation set. Lines and shaded areas show the average and standard deviation of 30 leave-simulation-out tests, and each panel shows the results for a different redshift. **Bottom panels.** Leave-redshift-out tests require optimizing 1 emulator with all measurements but the ones at a particular redshift, and then using measurements from this redshift as validation. Each panel shows the results of a different test.

leave-one-out tests may require extrapolating the emulator's predictions, and it is widely known that machine-learning methods do not extrapolate well.

In the top panels of Fig. 5, lines and shaded areas display the average and standard deviation of 30 leave-simulation-out tests. Each test requires optimizing an emulator with 29 distinct TRAINING simulations, and then using the remaining simulation as the validation. Each panel shows the results for a different redshift, and we check that the results are similar for redshifts not shown. As we can see, the large-scale noise is similar for all TRAINING simulations; this is because they use the same initial distribution of Fourier phases. The overall performance of FORESTFLOW in recovering $b_\delta$ and $b_\eta$ is 1.0 and 3.1%, respectively, which translates into 2.0 and 2.9% for perpendicular and parallel $P_{3D}$ modes on linear scales, and 3.4 and 1.8% for $P_{3D}$ and $P_{1D}$.

In Table 1, we gather the precision of FORESTFLOW at the center and across the parameter space, as well as the expected level of uncertainties due to cosmic variance and the limited flexibility of the $P_{3D}$ model. Given that we use our simulations to evaluate the precision of FORESTFLOW, and the impact of cosmic variance on simulation measurements is 1.3 and 0.5% on $P_{3D}$ and $P_{1D}$, respectively, these numbers limits the maximum precision we can test with our simulations. These levels would decrease by evaluating the precision using bigger simulations with the same resolution. Uncertainties due to the impact of cosmic variance on the training data and inaccuracies of the $P_{3D}$ model in describing $P_{3D}$ and $P_{1D}$ are 2.4 and 0.6% for these statistics, respectively, which is 1.1 and 0.1% worse than the impact of cosmic variance on simulation measurements. At the center of the parameter space, the precision of FORESTFLOW for $P_{3D}$ and $P_{1D}$

**Table 1.** Percent precision of the $P_{3D}$ model (Eqs. 3 and 4) and FORESTFLOW in recovering $P_{3D}$ and $P_{1D}$, as well as the impact of cosmic variance on these statistics. The second and third columns show the results for $P_{3D}$ and $P_{1D}$ over the intervals $0.5 < k[\,\mathrm{Mpc}^{-1}] < 5$ and $0.09 < k_{\parallel}[\,\mathrm{Mpc}^{-1}] < 4$ intervals, respectively, while the last two columns do so for the perpendicular and parallel modes of $P_{3D}$ on linear scales.

| Type | $P_{3D}$ | $P_{1D}$ | $b_\delta$ | $b_\eta$ | $P^{\rm lin}_{3D,\perp}$ | $P^{\rm lin}_{3D,\parallel}$ |
|---|---|---|---|---|---|---|
| Cvar. fit[a] | 0.8 | 0.1 | 0.6 | 1.8 | 1.2 | 1.8 |
| Cvar. data[b] | 1.3 | 0.5 | – | – | – | – |
| Cvar. & model[c] | 2.4 | 0.6 | – | – | – | – |
| Emu. center[d] | 2.6 | 0.8 | 1.2 | 0.3 | 1.1 | 1.2 |
| Emu. overall[e] | 3.4 | 1.8 | 1.0 | 3.1 | 2.0 | 2.9 |

**Notes.** [(a)] Impact of cosmic variance on the best-fitting fitting $P_{3D}$ model to simulation measurements (see Appendix A). [(b)] Impact of cosmic variance on simulation measurements (see Appendix A). [(c)] Joint impact of cosmic variance and the limited flexibility of the $P_{3D}$ model (see §3.4). [(d)] Precision of FORESTFLOW at the center of the parameter space, estimated using the CENTRAL simulation (see §5.1). [(e)] Precision of FORESTFLOW across the parameter space, estimated via leave-simulation-out tests (see §5.1).

is only 0.2% worse than the previous levels, letting us conclude that the primary factors limiting the performance of FORESTFLOW at the center of the parameter space are the size of the training simulations and model inaccuracies.

The efficiency of FORESTFLOW across the parameter space is 1.2 and 1.0% worse than at the center for $P_{3D}$ and $P_{1D}$, respectively. Consequently, the precision of the emulator would likely improve by increasing the number of training simulations. However, leave-one-out tests significantly underestimate the precision at the edges of the training set, especially for a small number of simulations, because it often requires extrapolating the emulator's predictions. We can thus conclude that the quality of the training data, the precision of the model, and the number of training simulations have a similar impact on the performance of FORESTFLOW. Given that leave-simulation-out tests tend provide a lower bound for the emulator's performance, we conclude that the overall precision of FORESTFLOW in predicting $P_{3D}$ from linear scales to $k = 5\,\mathrm{Mpc}^{-1}$ is approximately 3%, and $\simeq 1.5\%$ for $P_{1D}$ down to $k_{\parallel} = 4\,\mathrm{Mpc}^{-1}$.

As discussed in §2.2, FORESTFLOW uses as input parameters measured from each snapshot of the training simulations rather than "traditional" cosmological parameters such as $\Omega_m$, $A_s$, or $H_0$. This strategy enables training FORESTFLOW without specifying the input redshift and making predictions for redshifts not present in the training set. To test this assumption, we carry out two leave-redshift-out tests. The first involves optimizing one emulator with all TRAINING measurements but the ones at $z = 2.5$, and then validating it with data from this redshift. For the second, we follow the same approach but using measurements at $z = 3.5$. We display the results of these tests in the bottom panels of Fig. 5. The precision of the emulator is similar for leave-redshift-out and leave-simulation-out tests, validating the approach mentioned above. We find similar results for leave-redshift-out tests at other redshifts.

### 5.2. Cosmologies and IGM histories outside the training set

In Fig. 6, we examine the precision of FORESTFLOW reproducing $P_{3D}$ and $P_{1D}$ measurements from simulations not included in the training set. Lines indicate the redshift average of the relative difference between model predictions and simulation measurements. The first two rows show the results for the CENTRAL and SEED simulations, whose only difference is their initial distribution of phases. Consequently, the predictions of FORESTFLOW are the same for both. As we can see, these simulations present a different large-scale pattern of fluctuations, signaling that are caused by cosmic variance. Once we ignore these, we find that the precision of FORESTFLOW is practically the same for both simulations. We can thus conclude that FORESTFLOW predictions are largely insensitive to the impact of cosmic variance on the training set.

In the third, fourth, and fifth rows of Fig. 6, we use the GROWTH, NEUTRINOS, and CURVED simulations to evaluate the precision of FORESTFLOW for three different scenarios not contemplated in the training set: different growth history, massive neutrinos, and curvature. As we can see, the precision of FORESTFLOW for all these simulations is approximately the same as for the CENTRAL simulation. These results support that using the small-scale amplitude and slope of the linear power spectrum to capture cosmological information enables setting precise constraints on growth histories and ΛCDM extensions not included in the training set (see also Pedersen et al. 2021; Pedersen et al. 2023; Cabayol-Garcia et al. 2023).

In the last row of Fig. 6, we examine the precision of FORESTFLOW for the REIONISATION simulation, which employs a He II reionization history significantly different from those used by the TRAINING simulations. The precision of the emulator for this and the CENTRAL simulation is similar, which is noteworthy given that the performance of $P_{1D}$ emulators for the REIONISATION is significantly worse than for the CENTRAL simulation (Cabayol-Garcia et al. 2023). The outstanding performance of FORESTFLOW is likely because the relationship between IGM physics and the parameters of the $P_{3D}$ model is more straightforward than with $P_{1D}$ variations.

## 6. Discussion

Cosmological analyses of the Lyman-$\alpha$ forest come in two flavors: one-dimensional studies focused on small, non-linear scales and three-dimensional analyses of large, linear scales. With FORESTFLOW, we can now consistently model Lyman-$\alpha$ correlations from nonlinear to linear scales, enabling a variety of promising analyses that we discuss next.

### 6.1. Connecting large-scale biases with small-scale physics

Small-scale Lyman-$\alpha$ analyses use emulators to predict $P_{1D}$ as a function of cosmology and IGM physics (e.g.; Cabayol-Garcia et al. 2023), while large-scale analyses use linear or perturbation theory models to predict $\xi_{3D}$ together with Lyman-$\alpha$ linear bias parameters that need to be marginalized over. FORESTFLOW provides a relationship between IGM physics and linear biases, enabling the use of $P_{1D}$ studies to inform three-dimensional analyses and vice versa.

We could use FORESTFLOW to set constraints on $b_\delta$ and $b_\eta$ by fitting $P_{1D}$ measurements, and then use these constraints as priors in three-dimensional studies. As a result, we would break degeneracies between Lyman-$\alpha$ linear bias parameters and cosmology, allowing us to measure the amplitude of linear density and velocity fluctuations, $\sigma_8(z)$ and $f\sigma_8(z)$, rather than $b_\delta\sigma_8$ and $b_\eta f\sigma_8$ like in traditional Lyman-$\alpha$ forest analyses. To illustrate this application, we proceed to compare measurements of $b_\delta$ and $\beta$ from BAO analyses with FORESTFLOW predictions based $P_{1D}$ analyses. The analysis of BAO in the Lyman-$\alpha$ forest from

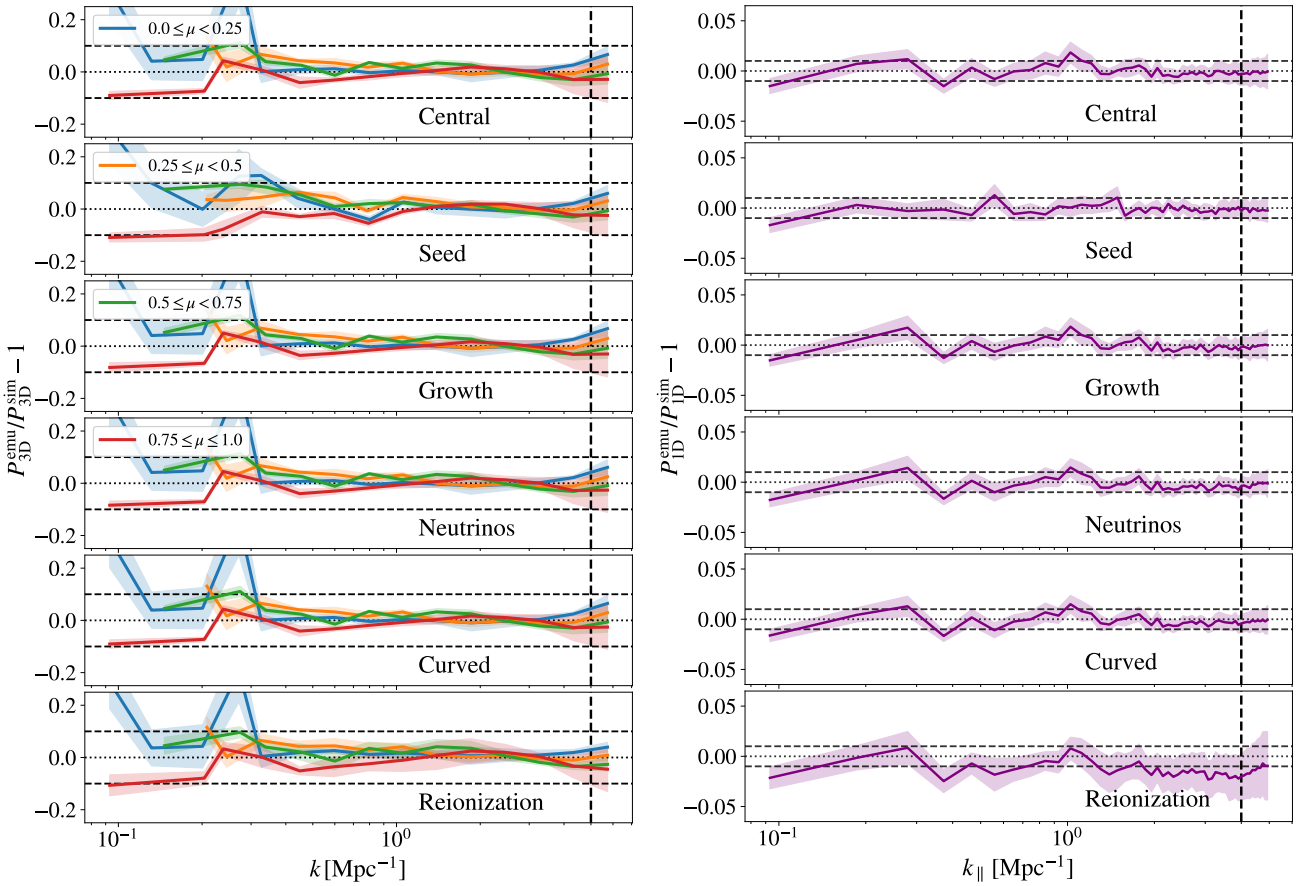


**Fig. 6.** Precision of the emulator in recovering $P_{3D}$ and $P_{1D}$ for test simulations not included in the training set. Lines and shaded areas display the average and standard deviation of the results for 11 snapshots between $z = 2$ and 4.5, respectively. From top to bottom, the rows show the results for the CENTRAL, SEED, GROWTH, NEUTRINOS, CURVED, and REIONISATION simulations, where the CENTRAL and SEED simulations are at the center of the input parameter space and employ the same and different initial distribution of Fourier phases as the training simulations, respectively, the GROWTH and REIONISATION simulations use a different growth and reionization history relative to those used by the TRAINING simulations, and the NEUTRINOS and CURVED simulations consider massive neutrinos and curvature. The efficiency of FORESTFLOW is approximately the same for all simulations.

the first data release of DESI yields $b_\delta = -0.108 \pm 0.005$ and $\beta = 1.74 \pm 0.09$ at $z = 2.33$ (DESI Collaboration et al. 2024). On the other hand, FORESTFLOW predicts $b_\delta = -0.118$ and $\beta = 1.57$ at $z = 2.33$ for a *Planck* cosmology when using as input the best-fitting constraints on IGM parameters from table 4 of Walther et al. (2019), which were derived from high-resolution $P_{1D}$ measurements. Even though the constraints on IGM parameters were derived using a $P_{1D}$ emulator trained on simulations with possibly slightly different definitions of IGM parameters relative to those used in this work, FORESTFLOW predictions and DESI measurements agree at the 2 sigma level, encouraging this new type of study.

In the left panels of Fig. 7, we display FORESTFLOW predictions for the response of the Lyman-$\alpha$ linear biases and $\beta$ to variations in cosmology and IGM physics. The response of $b_\eta$ to cosmology and IGM variations is quite weak, but the response of $b_\delta$ and $\beta$ to these changes is strong and redshift dependent. Therefore, we could analyze $P_{3D}$ measurements from different redshfits using FORESTFLOW to further break degeneracies between $b_\delta$ and $\sigma_8$. Note that the response of the Lyman-$\alpha$ linear biases and $\beta$ to $A_s$ variations broadly agrees with measurements from simulations run varying only one cosmological parameter at a time (Arinyo-i-Prats et al. 2015).

Similarly, we could use measurements of linear bias parameters from three-dimensional analyses (du Mas des Bourboux et al. 2020; DESI Collaboration et al. 2024) to make predictions for IGM parameters, which could be used in $P_{1D}$ studies to break degeneracies between cosmology and IGM physics. In the right panels of Fig. 7, we display FORESTFLOW predictions for the response of $P_{3D}$ and $P_{1D}$ to variations in cosmology and IGM physics. As we can see, the response of $P_{1D}$ to $A_s$ and $\bar{F}$ variations is largely scale-independent down to $k_\parallel = 1\,\mathrm{Mpc}^{-1}$ where many other effects are at play, and thus these two parameters are largely degenerated. On the other hand, this is not the case for $P_{3D}$, and thus we could use information from $P_{3D}$ analyses to break degeneracies in $P_{1D}$ studies. Note that the response of $P_{3D}$ and $P_{1D}$ to $A_s$, $\bar{F}$, and $\sigma_T$ variations broadly agrees with measurements from simulations run varying only one cosmological parameter at a time (McDonald 2003; McDonald et al. 2005).

We can also see $P_{3D}$ and $P_{1D}$ response strongly to variations in $\Omega_M h^2$. However, these variations can be largely absorbed by changes in the value of $A_s$ and $n_s$. Specifically, we check that changes in $A_s$ and $n_s$ completely absorb $\Omega_M h^2$ variations for $P_{3D}$ near the pivot scale we consider in our work $k_p = 0.7\,\mathrm{Mpc}^{-1}$, and progressively grow to 2% towards larger and smaller scales. For $P_{1D}$, we check that the same changes also absorb $\Omega_M h^2$ variations up to the 2% level. Nonetheless, it is important to note that *Planck* measured $\Omega_M h^2$ with 0.8% precision Planck Collaboration et al. (2020), and thus the precision of using the $\Delta_p^2\, n_p$ parametrization is actually much better for both $P_{3D}$ and $P_{1D}$.

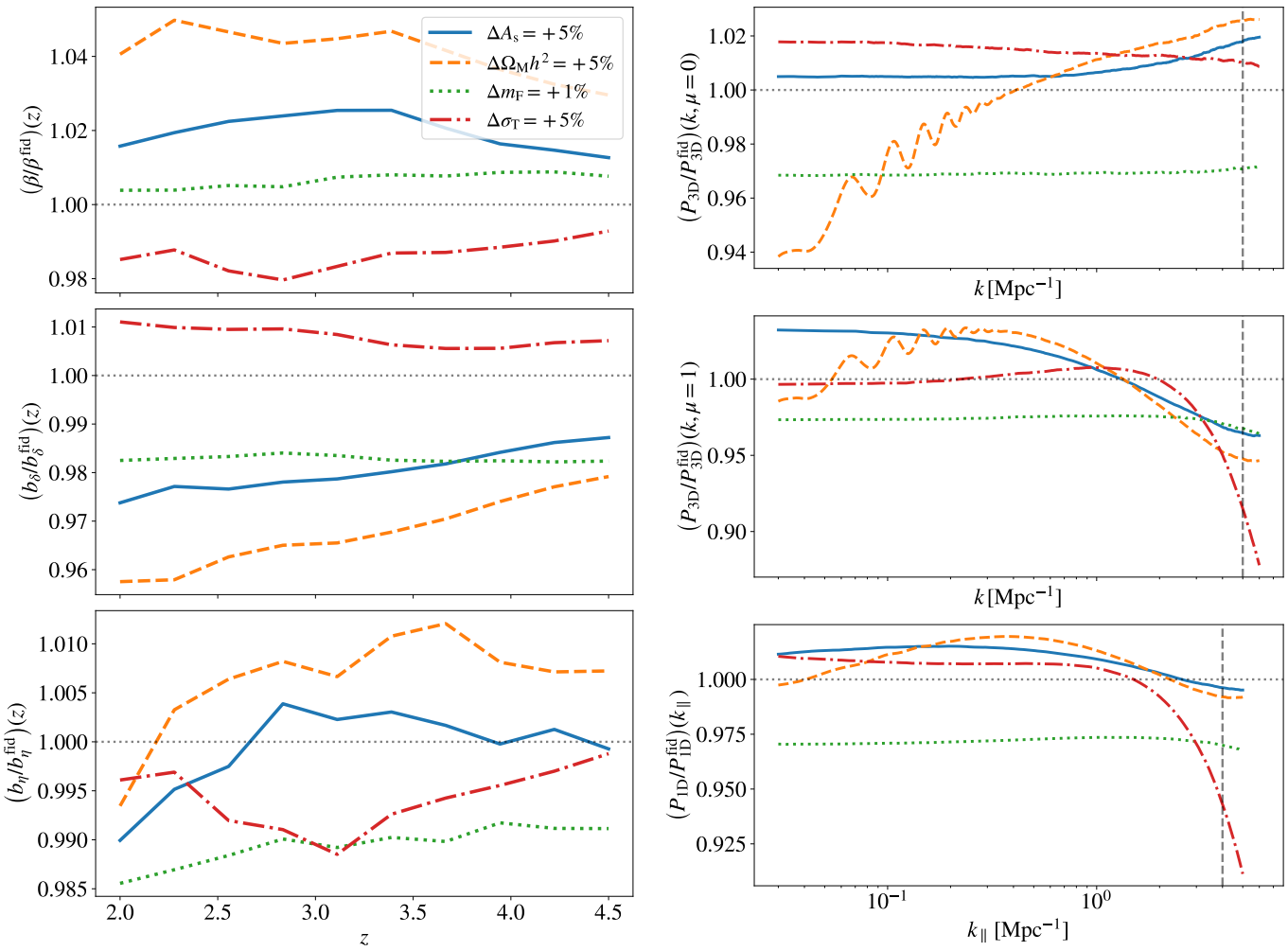


**Fig. 7.** Response of Lyman-$\alpha$ clustering to variations in cosmology and IGM physics according to FORESTFLOW. The top, middle, and bottom panels of the left column show the results for $\beta$, $b_\delta$, and $b_\eta$, respectively, while those of the right column do so for the perpendicular modes of $P_{3D}$, the parallel modes of $P_{3D}$, and $P_{1D}$. Blue, orange, and red lines show the response of the previous quantities to a 5% increase in $A_s$, $\Omega_M h^2$, and $\sigma_T$, respectively, while green lines do so for a 1% increase in $\bar{F}$.

### 6.2. Alcock-Paczyński on mildly non-linear scales

Thanks to the increasing precision of galaxy surveys, there is a growing interest in extracting cosmological information from increasingly smaller scales in three-dimensional analyses. An avenue to do so is to analyze anisotropies in the correlation function Alcock & Paczynski (AP test; 1979), first proposed in the context of the Lyman-$\alpha$ forest by McDonald & Miralda-Escudé (1999); Hui et al. (1999). Recently, Cuceu et al. (2023) followed this approach to analyze Lyman-$\alpha$ forest measurements from the Sloan Digital Sky Survey (SDSS) data release 16 (DR16; Ahumada et al. 2020), yielding constraints on some cosmological parameters a factor of two tighter than those from BAO-only analyses.

This study modeled three-dimensional correlations using linear theory, which restricted the range of scales analyzed to those larger than $25h^{-1}$ Mpc. We could significantly extend the range of scales used in this type of analysis by modeling three-dimensional correlations using FORESTFLOW. As a result, the constraining power of AP analyses would be much larger. Furthermore, we could use FORESTFLOW to extract information from $P_{1D}$ analyses to reduce degeneracies between cosmology and the parameters describing $\xi_{3D}$ (see §6.1).

### 6.3. Extending 3D analyses to the smallest scales

The ultimate goal of FORESTFLOW is to perform a joint analysis of one- and three-dimensional measurements from small to large scales. An interesting approach to do so is to measure the Lyman-$\alpha$ forest cross-spectrum ($P_\times$; e.g.; Hui et al. 1999; Font-Ribera et al. 2018), which captures the correlation between one-dimensional Fourier modes from two neighboring quasars separated by a transverse separation ($r_\perp$). We can model this statistic by taking the inverse Fourier transform of $P_{3D}$ only along the perpendicular directions

$$P_\times(k_\parallel, r_\perp) \equiv \frac{1}{(2\pi)^2}\int {\rm d}\boldsymbol{k}_\perp\, e^{i\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\, P_{3D}(k_\parallel, k_\perp)$$
$$= \frac{1}{2\pi}\int_0^\infty {\rm d}k_\perp\, k_\perp\, J_0(k_\perp r_\perp)\, P_{3D}(k_\parallel, k_\perp)\,. \qquad (11)$$

Comparing this equation with Eq. 1, it becomes clear that $P_{1D}$ is a special case of $P_\times$, corresponding to the limit where the transverse separation is zero.

In §3.3, we optimize the $P_{3D}$ model to describe measurements of $P_{3D}$ and $P_{1D}$ from the TRAINING simulations. Then, in §4, we use the distribution of best-fitting parameters as the training set for FORESTFLOW, which predicts the value of $P_{3D}$ model

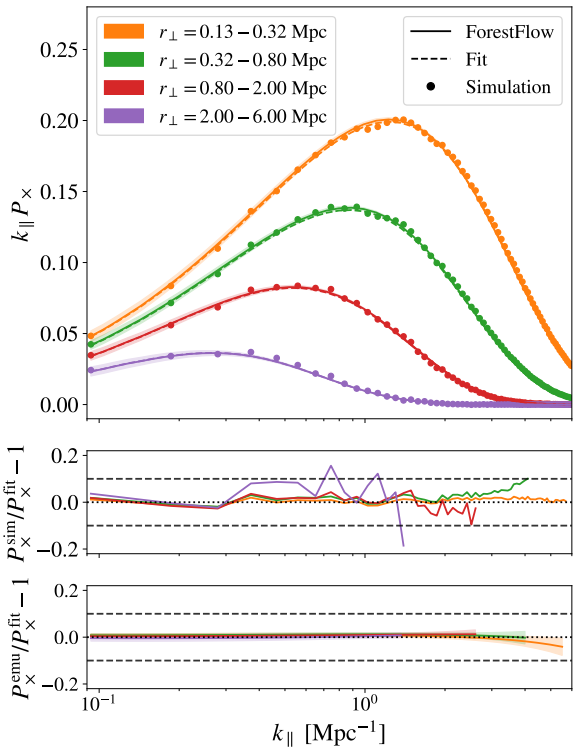


**Fig. 8.** Precision of the parametric model and the emulator in describing $P_\times$ measurements from the CENTRAL simulation at $z = 3$. Dots show simulation measurements, dashed lines depict predictions from the best-fitting parametric model to $P_{\rm 3D}$ and $P_{\rm 1D}$ measurements, and solid lines and shaded areas display the average and 68% credible interval of FORESTFLOW predictions. The color of the lines indicates the results for different bins in transverse separation $r_\perp$. The middle panel shows the residual between simulation measurements and model predictions, while the bottom panel displays the residual between model and emulator predictions. The precision of FORESTFLOW in reproducing simulation measurements is similar to that of the best-fitting model.

parameters as a function of cosmology and IGM physics. Even though neither the best-fitting model nor FORESTFLOW use $P_\times$ for their optimization, we can make predictions of $P_\times$ for both. To do so, we first estimate $P_{\rm 3D}$ using the value of the model parameters using Eq. 3, and then we integrate it using Eq. 11. We carry out the integration using the fast Hankel transform algorithm `FFTlog` (Hamilton 2000) implemented in the `hankl` package (Karamanis & Beutler 2021).

We use $P_\times$ measurements from the simulations described in § 3.1 to evaluate the precision of FORESTFLOW for this statistic. We first define four bins in $r_\perp$, the transverse separation between skewers in configuration space, with edges 0.13, 0.32, 0.80, 2, and 6 Mpc. Then, we measure $P_\times$ using all pairs of skewers with $r_\perp$ separation within the previous bins

$$P_\times(r_\perp, k_\parallel) = \left\langle \Re\left[\tilde{\delta}_{\rm i}(k_\parallel)\tilde{\delta}_{\rm j}^{\,*}(k_\parallel)\right]\right\rangle \tag{12}$$

where $\tilde{\delta}_{\rm i}$ and $\tilde{\delta}_{\rm j}^{\,*}$ stand for the Fourier transform of a skewer $i$ and the complex conjugate of its partner $j$, respectively, the average $\langle\rangle$ includes all possible pairs in the bin without repetition or permutation, and $\Re$ indicates that we only use the real part of the expression between brackets because the average of the imaginary part is zero. The $r_\perp$ on the left-hand side denotes the effective center of the bin, accounting for the skewed distribution of $r_\perp$ within each bin: the number of skewers separated by a small distance $dr_\perp$ is proportional to $r_\perp$, and therefore the effective center is not at the halfway point. To compute $P_\times$ at the effective center, we perform the integration using ten sub-bins within each $r_\perp$ bin and calculate the average of these weighted by $r_\perp$.

In Fig. 8, we study the precision of FORESTFLOW in reproducing $P_\times$ measurements from the CENTRAL simulation at $z = 3$. Dots display simulation measurements, dashed lines the best-fitting model to $P_{\rm 3D}$ and $P_{\rm 1D}$ measurements from this simulation, and the solid lines FORESTFLOW predictions. As we can see, $P_\times$ decreases as the $r_\perp$ separation increases; this is because more distant sightlines are sampling increasingly uncorrelated regions. In the middle panel, we examine the precision of the best-fitting model in describing simulation measurements, finding that it is better than 10% throughout all the scales shown. The precision of the model improves for smaller $r_\perp$ separations. This is likely because the fit's likelihood function (Eq. 7) considers $P_{\rm 1D}$, which is equivalent to $P_\times$ at $r_\perp = 0$ separation, but not $P_\times$. The bottom panel addresses the performance of FORESTFLOW relative to the best-fitting model; in this way, we approximately evaluate the performance of the emulator in reproducing the training data. The precision of FORESTFLOW in recovering the best-fitting model is better than 5% for all scales shown.

Future studies could use FORESTFLOW for extracting constraints on cosmology and IGM physics from the analysis of $P_\times$ measurements (e.g.; Abdul Karim et al. 2024). Nevertheless, as with $P_{\rm 1D}$, these analyses would also require modeling multiple systematics affecting Lyman-$\alpha$ measurements such as damped Lyman-$\alpha$ systems, metal line contamination, and AGN feedback.

## 7. Conclusions

We present FORESTFLOW, a cosmological emulator that predicts Lyman-$\alpha$ clustering from linear to nonlinear scales. Using an architecture based on conditional normalizing flows, FORESTFLOW emulates the 2 linear Lyman-$\alpha$ biases ($b_\delta$ and $b_\eta$) and 6 physically-motivated parameters capturing small-scale deviations of the three-dimensional flux power spectrum ($P_{\rm 3D}$) from linear theory. We summarize the main results of this work below:

- The main advantage of our strategy, compared to emulating $P_{\rm 3D}$ at a set of $k$-bins, is that FORESTFLOW can predict Lyman-$\alpha$ clustering on arbitrarily large (linear) scales when combined with a Boltzmann solver. Additionally, the emulator can make predictions for any statistics derived from $P_{\rm 3D}$ without interpolation or extrapolation, including the two-point correlation function ($\xi_{\rm 3D}$, main statistic of large-scale studies), the one-dimensional Lyman-$\alpha$ flux power spectrum ($P_{\rm 1D}$, main statistic of small-scale studies), and the cross-spectrum ($P_\times$, promising statistic for full-scale studies).
- To train the emulator, we use the best-fitting value of the 8 model parameters to $P_{\rm 3D}$ and $P_{\rm 1D}$ measurements from a suite of 30 fixed-and-paired cosmological hydrodynamical simulations spanning 11 equally-spaced redshifts between $z = 2$ and 4.5. We emulate these parameters as a function of the small-scale amplitude and slope of the linear power spectrum, the mean transmitted flux fraction, the amplitude and slope of the temperature-density relation, and the pressure smoothing scale (see Pedersen et al. 2021). We use this parameterization because it has the potential for making predictions for extensions to the $\Lambda$CDM model and ionization histories not included in the training set (Pedersen et al. 2023; Cabayol-Garcia et al. 2023).

- In §5.1, we show that the precision of FORESTFLOW in predicting $P_{3D}$ from linear scales to $k = 5\,\mathrm{Mpc}^{-1}$ is 3% and 1.5% for $P_{1D}$ down to $k_{\parallel} = 4\,\mathrm{Mpc}^{-1}$. We find that the size and number of training simulations have a similar impact on the emulator's performance as uncertainties arising from the limited flexibility of the 8-parameter model.
- In §5.2, we show that FORESTFLOW displays similar performance as before for two extensions to the ΛCDM model — massive neutrinos and curvature — and ionization histories not included in the training set.

The release of FORESTFLOW is timely for Lyman-$\alpha$ forest analyses with the ongoing Dark Energy Spectroscopic Instrument (DESI) survey. As noted in §6, FORESTFLOW enables a series of novel multiscale studies with DESI data, including connecting large- and small-scale analyses as well as extending three-dimensional analyses towards smaller scales.

*Acknowledgements.* JCM, LCG, ML, and AFR acknowledge support from the European Union's Horizon Europe research and innovation programme (COSMOLYA, grant agreement 101044612). AFR acknowledges financial support from the Spanish Ministry of Science and Innovation under the Ramon y Cajal program (RYC-2018-025210) and the PGC2021-123012NB-C41 project. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. The analysis has been performed at Port d'Informació Científica (PIC); we acknowledge the support provided by PIC in granting us access to their computing resources. This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of High-Energy Physics, under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Humanities, Science and Technology of Mexico (CONAHCYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: https://www.desi.lbl.gov/collaborating-institutions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U. S. National Science Foundation, the U. S. Department of Energy, or any of the listed funding agencies. The authors are honored to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

## Data Availability

FORESTFLOW and all the notebooks used to generate the plots of this paper can be found in https://github.com/igmhub/ForestFlow, as well as all data points shown in the published graphs. The simulations utilized for training and testing the emulator are publicly accessible at https://github.com/igmhub/LaCE.

## References

Abdul Karim, M. L., Armengaud, E., Mention, G., et al. 2024, J. Cosmology Astropart. Phys., 2024, 088
Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, ApJS, 249, 3
Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. 2019, arXiv e-prints, arXiv:1907.10902
Alcock, C. & Paczynski, B. 1979, Nature, 281, 358
Anderson, L., Pontzen, A., Font-Ribera, A., et al. 2019, ApJ, 871, 144
Angulo, R. E., Baugh, C. M., Frenk, C. S., & Lacey, C. G. 2008, MNRAS, 383, 755
Angulo, R. E. & Pontzen, A. 2016, MNRAS, 462, L1
Ansel, J., Yang, E., He, H., Gimelshein, N., & Jain, A. 2024, in ASPLOS 24(ACM)
Ardizzone, L., Bungert, T., Draxler, F., et al. 2018-2022, Framework for Easily Invertible Architectures (FrEIA)
Arinyo-i-Prats, A., Miralda-Escudé, J., Viel, M., & Cen, R. 2015, JCAP, 2015, 017
Bird, S. 2017, Astrophysics Source Code Library, ascl:1710.012
Bird, S., Feng, Y., Pedersen, C., & Font-Ribera, A. 2020, J. Cosmology Astropart. Phys., 2020, 002
Bird, S., Fernandez, M., Ho, M.-F., et al. 2023, JCAP, 2023, 037
Bird, S., Rogers, K. K., Peiris, H. V., et al. 2019, J. Cosmology Astropart. Phys., 2019, 050
Boera, E., Becker, G. D., Bolton, J. S., & Nasir, F. 2019, ApJ, 872, 101
Bolton, J. S., Puchwein, E., Sijacki, D., et al. 2017, MNRAS, 464, 897
Bolton, J. S., Viel, M., Kim, T. S., Haehnelt, M. G., & Carswell, R. F. 2008, MNRAS, 386, 1131
Busca, N. G., Delubac, T., Rich, J., et al. 2013, A&A, 552, A96
Cabayol-Garcia, L., Chaves-Montero, J., Font-Ribera, A., & Pedersen, C. 2023, MNRAS, 525, 3499
Cen, R., Miralda-Escudé, J., Ostriker, J. P., & Rauch, M. 1994, ApJ, 437, L9
Chabanier, S., Bournaud, F., Dubois, Y., et al. 2020, MNRAS, 495, 1825
Chabanier, S., Emberson, J. D., Lukić, Z., et al. 2023, MNRAS, 518, 3754
Chen, S.-F., Vlah, Z., & White, M. 2021, J. Cosmology Astropart. Phys., 2021, 053
Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1998, ApJ, 495, 44
Cuceu, A., Font-Ribera, A., Nadathur, S., Joachimi, B., & Martini, P. 2023, Phys. Rev. Lett., 130, 191003
Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
de Belsunce, R., Philcox, O. H. E., Irsic, V., et al. 2024, arXiv e-prints, arXiv:2403.08241
DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2024, arXiv e-prints, arXiv:2404.03001
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv e-prints, arXiv:1611.00036
Dinh, L., Sohl-Dickstein, J., & Bengio, S. 2016, arXiv e-prints, arXiv:1605.08803
du Mas des Bourboux, H., Rich, J., Font-Ribera, A., et al. 2020, ApJ, 901, 153
Feng, Y., Bird, S., Anderson, L., Font-Ribera, A., & Pedersen, C. 2018, Zenodo
Fernandez, M. A., Ho, M.-F., & Bird, S. 2022, MNRAS, 517, 3200
Font-Ribera, A., McDonald, P., & Slosar, A. 2018, JCAP, 2018, 003
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gaikwad, P., Rauch, M., Haehnelt, M. G., et al. 2020, MNRAS, 494, 5091
Gaikwad, P., Srianand, R., Haehnelt, M. G., & Choudhury, T. R. 2021, MNRAS, 506, 4389
Gerardi, F., Cuceu, A., Font-Ribera, A., Joachimi, B., & Lemos, P. 2023, MNRAS, 518, 2567
Givans, J. J., Font-Ribera, A., Slosar, A., et al. 2022, JCAP, 2022, 070
Givans, J. J. & Hirata, C. M. 2020, Phys. Rev. D, 102, 023515
Gnedin, N. Y. & Hui, L. 1998, MNRAS, 296, 44
Haardt, F. & Madau, P. 2012, ApJ, 746, 125
Hamilton, A. J. S. 2000, MNRAS, 312, 257
Hastie, T., Tibshirani, R., & Friedman, J. 2001, The Elements of Statistical Learning, Springer Series in Statistics (New York, NY, USA: Springer New York Inc.)
Horowitz, B., de Belsunce, R., & Lukic, Z. 2024, arXiv e-prints, arXiv:2403.17294
Huang, C.-W., Krueger, D., Lacoste, A., & Courville, A. 2018, arXiv e-prints, arXiv:1804.00779
Hui, L., Stebbins, A., & Burles, S. 1999, ApJ, 511, L5
Iršič, V., Viel, M., Haehnelt, M. G., Bolton, J. S., & Becker, G. D. 2017, Phys. Rev. Lett., 119, 031302
Iršič, V. & McQuinn, M. 2018, J. Cosmology Astropart. Phys., 2018, 026
Iršič, V., Viel, M., Haehnelt, M. G., et al. 2024, Phys. Rev. D, 109, 043511
Ivanov, M. M. 2024, Phys. Rev. D, 109, 023507
Jimenez Rezende, D. & Mohamed, S. 2015, arXiv e-prints, arXiv:1505.05770
Kaiser, N. 1987, MNRAS, 227, 1
Karaçaylı, N. G., Martini, P., Guy, J., et al. 2024, MNRAS, 528, 3941
Karamanis, M. & Beutler, F. 2021, arXiv e-prints, arXiv:2106.06331
Khan, N. K., Kulkarni, G., Bolton, J. S., et al. 2024, MNRAS, 530, 4920
Kingma, D. P. & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980
Kokron, N., Chen, S.-F., White, M., DeRose, J., & Maus, M. 2022, J. Cosmology Astropart. Phys., 2022, 059
Lee, K.-G., Hennawi, J. F., Spergel, D. N., et al. 2015, ApJ, 799, 196
Lesgourgues, J. 2011, arXiv e-prints, arXiv:1104.2932
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Lukić, Z., Stark, C. W., Nugent, P., et al. 2015, MNRAS, 446, 3697
MacKay, D. J. et al. 1998, NATO ASI series F computer and systems sciences, 168, 133
Maion, F., Angulo, R. E., & Zennaro, M. 2022, J. Cosmology Astropart. Phys., 2022, 036
McCulloch, W. S. & Pitts, W. 1943, The bulletin of mathematical biophysics, 5, 115
McDonald, P. 2003, ApJ, 585, 34

McDonald, P. & Miralda-Escudé, J. 1999, ApJ, 518, 24
McDonald, P., Miralda-Escudé, J., Rauch, M., et al. 2000, ApJ, 543, 1
McDonald, P., Seljak, U., Burles, S., et al. 2006, ApJS, 163, 80
McDonald, P., Seljak, U., Cen, R., et al. 2005, ApJ, 635, 761
McKay, M. D., Beckman, R. J., & Conover, W. J. 1979, Technometrics, 21, 239
McQuinn, M. 2016, ARA&A, 54, 313
Meiksin, A., Bryan, G., & Machacek, M. 2001, MNRAS, 327, 296
Meiksin, A. A. 2009, Reviews of Modern Physics, 81, 1405
Miralda-Escudé, J., Cen, R., Ostriker, J. P., & Rauch, M. 1996, ApJ, 471, 582
Molaro, M., Iršič, V., Bolton, J. S., et al. 2023, MNRAS, 521, 1489
Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2017, ApJ, 837, 106
Palanque-Delabrouille, N., Yèche, C., Lesgourgues, J., et al. 2015, JCAP, 2015, 045
Palanque-Delabrouille, N., Yèche, C., Schöneberg, N., et al. 2020, JCAP, 2020, 038
Papamakarios, G., Nalisnick, E., Jimenez Rezende, D., Mohamed, S., & Lakshminarayanan, B. 2019, arXiv e-prints, arXiv:1912.02762
Pedersen, C., Font-Ribera, A., & Gnedin, N. Y. 2023, ApJ, 944, 223
Pedersen, C., Font-Ribera, A., Rogers, K. K., et al. 2021, J. Cosmology Astropart. Phys., 2021, 033
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Pontzen, A., Slosar, A., Roth, N., & Peiris, H. V. 2016, Physical Review D, 93, 103519
Puchwein, E., Bolton, J. S., Keating, L. C., et al. 2023, MNRAS, 519, 6162
Puchwein, E., Haardt, F., Haehnelt, M. G., & Madau, P. 2019, MNRAS, 485, 47
Ramachandra, N., Chaves-Montero, J., Alarcon, A., et al. 2022, MNRAS, 515, 1927
Ravoux, C., Abdul Karim, M. L., Armengaud, E., et al. 2023, MNRAS, 526, 5118
Rogers, K. K. & Peiris, H. V. 2021a, Phys. Rev. D, 103, 043526
Rogers, K. K. & Peiris, H. V. 2021b, Phys. Rev. Lett., 126, 071302
Rogers, K. K., Peiris, H. V., Pontzen, A., et al. 2019, J. Cosmology Astropart. Phys., 2019, 031
Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. 1989, Statistical Science, 4, 409
Seljak, U., Makarov, A., McDonald, P., et al. 2005, Physical Review D, 71, 103515
Seljak, U., Makarov, A., McDonald, P., & Trac, H. 2006, Phys. Rev. Lett., 97, 191303
Seljak, U., Slosar, A., & McDonald, P. 2006, JCAP, 2006, 014
Slosar, A., Font-Ribera, A., Pieri, M. M., et al. 2011, JCAP, 2011, 001
Slosar, A., Iršič, V., Kirkby, D., et al. 2013, JCAP, 2013, 026
Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJS, 148, 175
Springel, V. 2005, MNRAS, 364, 1105
Takhtaganov, T., Lukić, Z., Müller, J., & Morozov, D. 2021, ApJ, 906, 74
Verde, L., Peiris, H. V., Spergel, D. N., et al. 2003, ApJS, 148, 195
Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, Phys. Rev. D, 88, 043502
Viel, M. & Haehnelt, M. G. 2006, MNRAS, 365, 231
Viel, M., Weller, J., & Haehnelt, M. G. 2004, MNRAS, 355, L23
Villaescusa-Navarro, F., Naess, S., Genel, S., et al. 2018, ApJ, 867, 137
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Walther, M., Armengaud, E., Ravoux, C., et al. 2021, J. Cosmology Astropart. Phys., 2021, 059
Walther, M., Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2019, ApJ, 872, 13
Winkler, C., Worrall, D., Hoogeboom, E., & Welling, M. 2019, arXiv e-prints, arXiv:1912.00042
Zaldarriaga, M., Hui, L., & Tegmark, M. 2001, ApJ, 557, 519

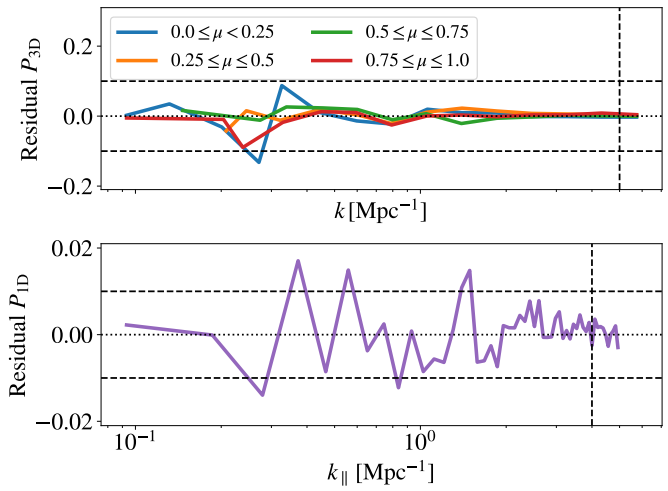


**Fig. A.1.** Impact of cosmic variance on $P_{3D}$ (top panel) and $P_{1D}$ (bottom panel) measurements from our simulations at $z = 3$. Lines show the difference between measurements from the CENTRAL and SEED simulations, which only differ on their initial distribution of Fourier phases, divided by $\sqrt{2}$ times their average. Cosmic variance induces errors as large as 10% on $P_{3D}$ for $k \simeq 0.3\,\mathrm{Mpc}^{-1}$, while these are of the order of 1% for $P_{1D}$.

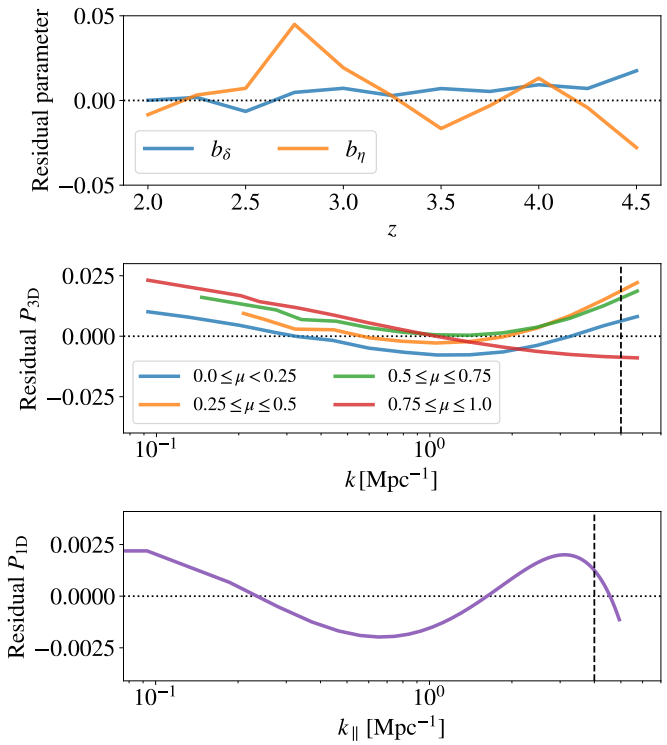


**Fig. A.2.** Impact of cosmic variance on predictions from the $P_{3D}$ parametric model. Lines show the difference between the best-fitting models to $P_{3D}$ and $P_{1D}$ measurements from the CENTRAL and SEED simulations, divided by $\sqrt{2}$ times the best-fitting model to their average. The top panel shows the results for the Lyman-$\alpha$ linear biases ($b_\delta$ and $b_\eta$), while the middle and bottom panels display the results for $P_{3D}$ and $P_{1D}$ at $z = 3$, respectively. The impact of cosmic variance on model predictions is approximately an order of magnitude smaller than on simulation measurements (see Fig. A.1).

## Appendix A: Cosmic variance

Throughout this work, we train and test FORESTFLOW using simulations run employing the "fixed-and-paired" technique (Angulo & Pontzen 2016; Pontzen et al. 2016), which significantly reduces cosmic variance for the clustering of the Lyman-$\alpha$ forest (Anderson et al. 2019). We could further mitigate the impact of cosmic variance by using control variates (Kokron et al. 2022), but this is outside the scope of the current work. The impact of cosmic variance on fixed-and-paired simulations is not straightforward (Maion et al. 2022), and thus we would ideally use multiple fixed-and-paired simulations with different initial distributions of Fourier phases to estimate the precision of simulation measurements. However, we only have two simulations with these properties: CENTRAL and SEED. In this section, we use these two simulations to estimate the impact of cosmic variance on simulation measurements and best-fitting models.

In Fig. A.1, we display the difference between measurements from the CENTRAL and SEED at $z = 3$ divided by $\sqrt{2}$ times their average[7]. The CENTRAL and SEED simulations only differ on their initial distribution of Fourier phases, and thus their difference isolates the impact of cosmic variance. In contrast to traditional simulations, where cosmic variance is inversely proportional to the square root of the number of modes for $P_{3D}$, this source of uncertainty reaches $\simeq 10\%$ at $k \simeq 0.3\,\mathrm{Mpc}^{-1}$ and decreases at both larger and smaller scales. This trend can be explained as follows: the reduction at the largest scales is due to the fix-and-paired technique, which completely cancels out cosmic variance for linear density modes. Conversely, the increase at intermediate scales is attributed to non-linear evolution, particularly mode coupling, which reintroduces cosmic variance on mildly non-linear scales. For even smaller scales, the number of modes increases, leading to a decrease in cosmic variance, similar to what is observed in traditional simulations.

The main consequence of cosmic variance is that it hinders our ability to evaluate the precision of the model and the emulator with our simulations. To mitigate the impact of this source of uncertainty, we quote the precision of the model and emulator for $P_{3D}$ on scales smaller than $k = 0.5\,\mathrm{Mpc}^{-1}$ throughout the main body of the test. Conversely, the impact of cosmic variance on $P_{1D}$ is approximately 1.5% at $k_\parallel < 2\,\mathrm{Mpc}^{-1}$, much smaller than for $P_{3D}$, letting us include all scales in the tests without concern.

For a more precise estimation, we compute the standard deviation of the results shown in Fig. A.1 across redshift. We do so within the intervals $0.5 < k\,[\mathrm{Mpc}^{-1}] < 5$ and $0.09 < k_\parallel\,[\mathrm{Mpc}^{-1}] < 4$ for $P_{3D}$ and $P_{1D}$, respectively, motivated by the previous discussion and the range of scales used when fitting the $P_{3D}$ model in §3.3. We find that the average impact of cosmic variance on $P_{3D}$ and $P_{1D}$ is 1.3 and 0.5%, respectively.

We now proceed to study the impact of cosmic variance on the best-fitting model for simulation measurements of $P_{3D}$ and $P_{1D}$. We anticipate that the impact of cosmic variance on the best-fitting model will be weaker than on individual simulation measurements because multiple $P_{3D}$ and $P_{1D}$ bins collectively contribute to determining the values of the 8 model parameters. In Fig. A.2, we show the difference between the best-fitting model to the CENTRAL and SEED simulations, divided by the $\sqrt{2}$ times the best-fitting model to their average. In the top panel, we show the results for the 2 Lyman-$\alpha$ linear biases ($b_\delta$ and $b_\eta$). The standard deviation of the differences is 0.6 and 1.8% for $b_\delta$ and $b_\eta$, respectively, and thus we can measure the 2 Lyman-$\alpha$ linear biases

[7] We use the factor $\sqrt{2}$ to estimate the noise for a single simulation.

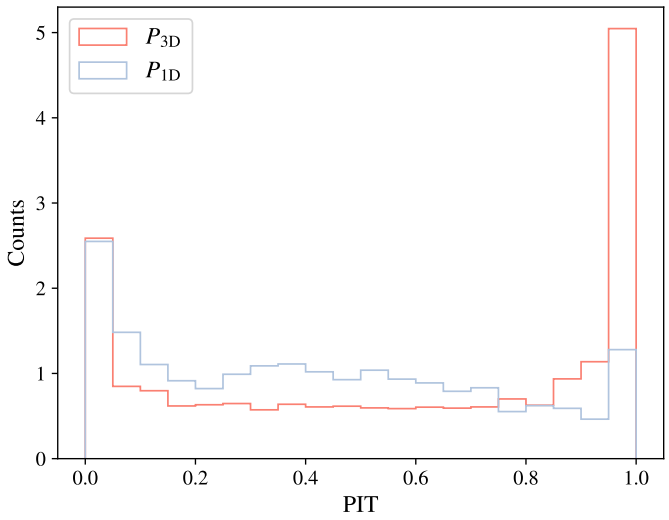


**Fig. B.1.** PIT distribution for $P_{1D}$ (blue) and $P_{3D}$ (red). This plot validates the uncertainties predicted by FORESTFLOW across TRAINING simulations via a leave-simulation-out approach. The PIT distribution is approximately uniform, indicating well-calibrated uncertainties for most samples, while the peaks at the edges indicate underestimated uncertainties for some samples.

with percent level precision from our simulations. By propagating these uncertainties to the behavior of $P_{3D}$ on linear scales, we find that the impact of cosmic variance on perpendicular and parallel modes is 1.2 and 1.8%, respectively.

In the middle and bottom panels of Fig. A.2, we address the influence of cosmic variance on model predictions for $P_{3D}$ and $P_{1D}$, respectively. The overall impact of this source of error on $P_{3D}$ and $P_{1D}$ is 0.8 and 0.1%, respectively, confirming that the best-fitting model is less sensitive to cosmic variance than simulation measurements. Consequently, FORESTFLOW is more robust against this type of uncertainty than emulators predicting the power spectrum at a set of $k$-bins.

## Appendix B: $P_{3D}$ and $P_{1D}$ uncertainty validation

Normalizing flows predict the full posterior distribution of the target data rather than only their mean like fully-connected neural networks or their mean and width like Mixture Density Networks (see Ramachandra et al. 2022; Cabayol-Garcia et al. 2023, for some applications in cosmology). This is achieved through multiple sampling iterations from the target latent distribution, an 8-dimensional Gaussian in our case. In FORESTFLOW, each sampled realization of the $P_{3D}$ model parameters is propagated to generate predictions for $P_{3D}$ and $P_{1D}$ (see §4.1), producing a covariance matrix for these statistics. In this appendix, we validate its diagonal elements. Note that well-calibrated uncertainties are critical for future uses of the emulator such as cosmology inference.

We validate the uncertainty in $P_{3D}$ and $P_{1D}$ predictions using the Probability Integral Transform test (PIT), which is the value of the cumulative distribution function (CDF) of a distribution evaluated at the ground-truth value $z_t$

$$\mathrm{PIT} = \mathrm{CDF}[\mathrm{p},\, \mathrm{z_t}] = \int_{-\infty}^{\mathrm{z_t}} \mathrm{p(z)dz}\,, \tag{B.1}$$

where $p$ is in our case the distribution of FORESTFLOW predictions for $P_{3D}$ or $P_{1D}$ and $z_t$ stands for measurements of these statistics from the simulations. A model that displays a well-calibrated uncertainty distribution yields PIT values that are uniformly distributed between zero and one. This indicates that the observed outcomes have an equal likelihood of falling at any point along the predicted CDF. In contrast, an excess of values close to zero or one indicates that the width of the distribution is underestimated.

In Fig. B.1, we display a PIT test produced using all the TRAINING simulations via a leave-simulation-out approach (see §5.1). This process validates average predictions and uncertainties against simulations excluded in the training process. The red and blue lines display the results for $P_{3D}$ and $P_{1D}$, respectively, which were generated by combining results from different scales and redshifts. As we can see, the PIT distribution is approximately uniform for both statistics but it presents peaks at the low and high ends, indicating underestimated uncertainties for some samples. The cause behind this feature is unclear and it demands further investigation beyond the scope of this project.