Prepared for submission to JCAP

# Parametrisation of a Lyman-$\alpha$ forest emulator.

**Chris Pedersen**[a]

[a]Department of Physics and Astronomy, University College London, Gower Street, London, United Kingdom

E-mail: chrisp@star.ucl.ac.uk

**Abstract.** Set of notes discussing parametrisations of an emulator for cosmology from the Lyman-$\alpha$ forest. This is meant primarily as an intituive introduction to our parametrisation rather than a discussion of the technical details, some of which will be included in Appendix C.

## Contents



# 1 Introduction

Constraining parameters of a cosmological model requires a large number of theoretical predictions to compare to observational data. In the case of the Lyman-$\alpha$ (Ly$\alpha$) forest, generating these theoretical predictions are computationally expensive which makes a brute-force approach unfeasible. Interpolation schemes have therefore been proposed where theoretical predictions are made at certain points in parameter space, and these then are used to infer the observable quantities for an arbitrary model. Our fundamental assumption is that the flux power spectrum of the Ly$\alpha$ forest at any given redshift is determined by the amplitude and shape of the linear matter power spectrum, and the state of the intergalactic medium (IGM). Under this assumption the standard $\Lambda$CDM parameters become impractical for Ly$\alpha$ forest analysis due to parameter degeneracies and difficulties in parametrising the state of the IGM. Therefore instead of constraining the cosmological parameters of the $\Lambda$CDM model directly, we intend to constrain the amplitude and slope of the linear power spectrum and its evolution with redshift, while marginalising over the IGM parameters (an approach that was taken in ref [1]). For our purposes we then need to introduce three different parameter spaces:

- an *emulator space* where we train a Gaussian process to provide a predicted 1D flux power spectrum for a given set of parameters that we believe to be closest to the data,
- a *likelihood space* which describes the amplitude and redshift evolution of the linear matter power spectrum and the IGM parameters, which is where we run our sampler,
- a *simulation space*, which consists of the parameters which we can input into MP-Gadget simulations (these are mostly $\Lambda$CDM parameters with some parameters which affect the IGM history)

# 2 Parameter spaces

## 2.1 Emulator space

Due to the fact that the universe is very close to Einstein-de-Sitter in the relevant redshfit range ($5 > z > 2$), the growth factor, $D(z)$ and the evolution of the Hubble expansion rate, $H(z)$, are close to independent of cosmology. The Ly$\alpha$ forest is also sensitive to only a small range of length scales within this redshift range (approximately $0.1\text{Mpc}^{-1} < k < 10\text{Mpc}^{-1}$), meaning that the large scale shape of the power spectrum also has no influence on the Ly$\alpha$ forest . Given this, the effect of changing various $\Lambda$CDM parameters on the 1 dimensional flux power spectrum (P1D) of the Ly$\alpha$

forest are highly degenerate with one another. Many of these effects will also arise through changing the thermal history of the IGM, which is something one would ideally decouple from cosmology and marginalise over. Instead we describe each measured P1D by a set of parameters which describe the linear power spectrum and the thermal state of the IGM. We use the following parameters:

- $\Delta_p^2$: the amplitude of the matter power spectrum at $k = 0.7\text{Mpc}^{-1}$. This scale is chosen as it is both a linear mode, and a length scale which is measured well in the P1D of the Ly$\alpha$ forest .
- $n_p$: the slope of the matter power spectrum around $k = 0.7\text{Mpc}^{-1}$,
- $\langle F \rangle$: the mean flux measured along the lines of sight in a given simulation snapshot. This property is affected mainly by the intensity of the UV background and the recombination rate, as well as the amount of clustering.
- $\sigma_T$: the thermal broadening scale. This quantifies the effect of the instantaneous gas temperature on smoothing the small scale flux power by measuring the temperature at mean density.
- $\gamma$: the relation between gas temperature and density is well approximated by a power law, where the exponent to this power law is $\gamma - 1$.
- $k_F$: the pressure smoothing scale. While the flux power along the line of sight is dependent on the instantaneous gas temperature due to Doppler smoothing, there is also a history-dependent smoothing scale caused by gas pressure[2][3].

We define our emulator parameter space as $\Phi = \{\Delta_p^2, n_p, \langle F \rangle, \sigma_T, \gamma, k_F\}$, and we refer to each P1D with an associated $\Phi$ as a *model*. Note that we do not include redshift as a parameter, as we propose that two flux power spectra at different redshifts but with the same $\Phi$ would have indistinguishable flux power spectra at our current precision. This parametrisation can be split into two groups - parameters that describe the linear power, and parameters that describe the state of the IGM, which we will consider nuisance parameters. Recent observational constraints on these IGM properties are available in [4] and can be used to motivate our priors. We note that one can modify the particle properties of the simulation data in postprocessing to explore a wider range of $\langle F \rangle$ and $\sigma_T$ models without running extra simulations. In appendix A we show the the effect of changing each of these parameters on the P1D.

### 2.2 Likelihood space

The likelihood parameter space is where we run our sampler, and these are the parameters for which we ultimately get constraints. In the case of the nuisance parameters, these are simply power laws in the emulator parameters of the form:

$$f(z) = A\left(\frac{1+z}{1+z_\star}\right)^B \tag{2.1}$$

where we swap $f(z)$ for the relevant IGM parameter, with the pivot redshift $z_\star = 3$. We denote the likelihood parameters as $\Theta = \{\Delta_\star^2, n_\star, \tau_0, \tau_1, T_1, T_2, T_3, \gamma_0, \gamma_1, k_{F0}, k_{F1}\}$, where:

- $\Delta_\star^2$: which describes the amplitude of the linear power spectrum at the pivot scale and redshift.
- $n_\star$: the slope of the power spectrum at the pivot scale and redshift,
- $\tau_0$ and $\tau_1$, which equivalently describe the amplitude and slope of $\ln(\langle \text{F} \rangle)$ around $z = 3$,
- $T_1$, $T_2$, $T_3$ which describe the redshift evolution of the thermal broadening scale, $\sigma_T$. The effect of He II reionisation on the IGM means that a broken power law is required to describe the evolution of the gas temperature (see Appendix B). Currently the break is fixed at $z = 3.6$.
- $\gamma_0$, $\gamma_1$, and $k_{F0}$, $k_{F1}$ which equivalently describe the amplitude and evolution of $\gamma$ and $k_F$ with redshift.

Some experimentation will be required to determine the optimum number of parameters to use for each power law. The specific redshifts at which emulator calls are made are set by the data, and one advantage of our parametrisation is that this can be set arbitrarily instead of being decided in advance.

### 2.3 Simulation space

These parameters are used to generate initial conditions and run the MP-Gadget simulations. We denote them $\Gamma = \{A_s, n_s, \Omega_\lambda, \Omega_{\rm b}, \Omega_{\rm c}, h, z_{\rm reion}, H_a, H_b\}$. The first six parameters are the standard parameters of $\Lambda$CDM. The remaining parameters control the evolution of the IGM - $z_{\rm reion}$ sets the median redshfit of reionisation, and $H_a$ and $H_b$ set the amplitude and density dependence of the gas heating. These final two parameters have the strongest effect on $\sigma_T$ and $\gamma$ respectively. For each simulation snapshot it is trivial to calculate the associated emulator space parameter, $\Delta_p$, using CAMB. However the inverse mapping is not unique and also requires a fiducial cosmology.

### 2.4 Example: maximising a likelihood

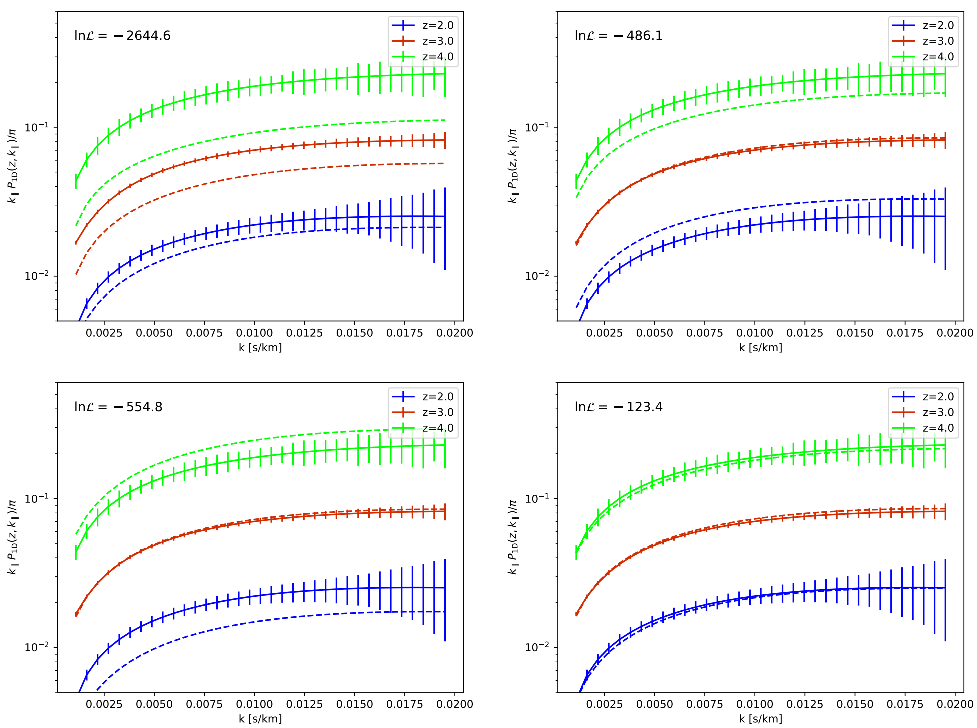


**Figure 1**. Predicted 1D flux power spectra (dashed) for 4 different sets of likelihood parameters, $\Theta$, overlaid on some mock data (solid lines with error bars) generated in MP-Gadget, using the covariance matrices from [5]. The top left panel shows a randomly chosen point in likelihood space. In the top right we show emulator calls for likelihood parameters with the correct $\tau_0$, but a low $\tau_1$, resulting in insufficient redshift evolution of $\langle F \rangle$. In the bottom left, we show another set of emulator calls for the correct $\tau_0$ but this time a $\tau_1$ that is too high. In the bottom right we show the emulator calls for the correct $\tau_0$ and $\tau_1$ which match the data well at all redshifts.

As an example likelihood maximisation, we consider a test case where we are only interested in constraining the likelihood parameters $\tau_0$ and $\tau_1$, with data redshift bins at $z = 4, 3, 2$. These parameters govern the amplitude and redshift evolution of $\langle F \rangle$. The effect of this parameter is essentially just a rescaling of the overall amplitude of the flux power spectrum, as can be seen in the top right panel of Appendix A. Initially $\tau_0$ and $\tau_1$ are randomly chosen from within the prior volume, leading to 3 emulator calls evaluating equation (2.1) at each redshift where we have data.

The likelihood evaluation is then the sum of the log likelihoods for each emulator call. This is shown in Fig. 1, where the mock data is shown in solid lines. In dashed lines we show the predicted flux power spectra for four different choices of $\tau_0$ and $\tau_1$. In the first case (top left), we show the predicted P1D for a randomly selected $\tau_0$ and $\tau_1$, which is clearly a poor fit to the data.
In the top right figure, we by hand fix $\tau_0$ to the correct value in the mock data, which results in the $z = 3$ power spectra matching the data well. The data at other redshifts is a poor match however as $\tau_1$ which governs the redshift evolution is far from the correct value. In the bottom left figure, we show emulator calls for likelihood parameters where $\tau_0$ is still correct, however $\tau_1$ is n ow far too high instead of too low. In this case there is too much evolution in the mean flux, $\langle F \rangle$, which means the overall amplitude of the P1D changes too drastically. In the final, bottom right panel, we show the emulator calls for the correct values of $\tau_0$ and $\tau_1$, where there is a strong match between the emulator predictions and the data at all redshifts. This is a particularly intuitive example because of the way in which the P1D depends on $\langle F \rangle$, but the principle can be extended to all other parameters and a larger number of redshift bins.

## 3 Comparison with previous approaches

(Could add some text here comparing with Nathalie's approach?)

## References


[1] P. McDonald, U. Seljak, R. Cen, D. Shih, D. H. Weinberg, S. Burles et al., *The Linear Theory Power Spectrum from the Ly$\alpha$ Forest in the Sloan Digital Sky Survey*, *ApJ* **635** (2005) 761 [astro-ph/0407377].

[2] L. Hui and N. Y. Gnedin, *Equation of state of the photoionized intergalactic medium*, Mon. Not. Roy. Astron. Soc. **292** (1997) 27 [astro-ph/9612232].

[3] N. Y. Gnedin and L. Hui, *Probing the Universe with the Lyalpha forest - I. Hydrodynamics of the low-density intergalactic medium*, Mon. Not. Roy. Astron. Soc. **296** (1998) 44 [astro-ph/9706219].

[4] M. Walther, J. Oñorbe, J. F. Hennawi and Z. Lukić, *New Constraints on IGM Thermal Evolution from the Ly$\alpha$ Forest Power Spectrum*, *ApJ* **872** (2019) 13 [1808.04367].

[5] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. Le Goff, G. Rossi, M. Viel et al., *The one-dimensional Ly$\alpha$ forest power spectrum from BOSS*, Astron. Astrophys. **559** (2013) A85 [1306.5896].

## A Parameter dependence of the 1-dimensional flux power spectrum

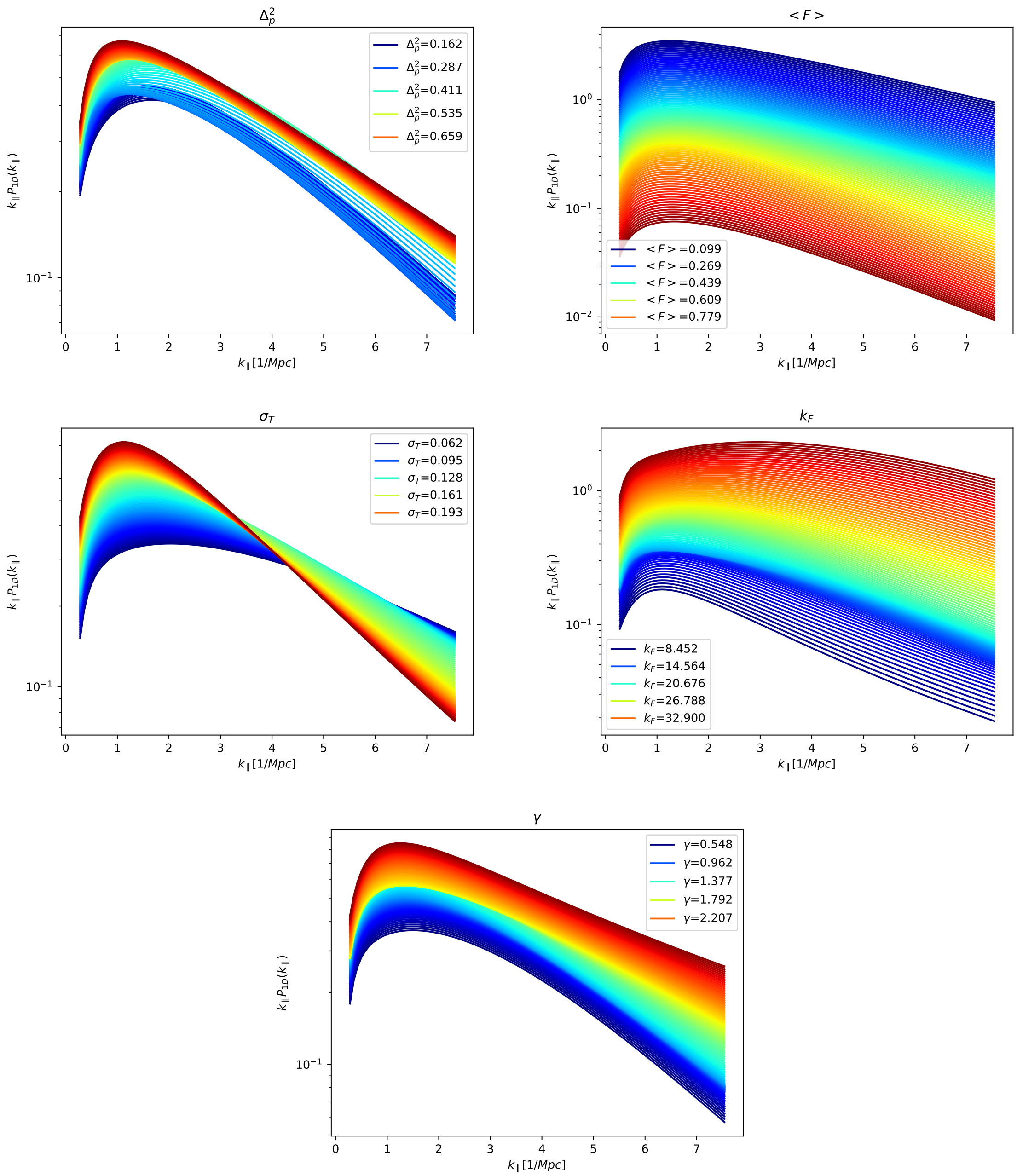


**Figure 2**. Dependence of the P1D on the different emulator parameters. Predictions are made using a Gaussian process emulator trained on $\sim$ 6000 models. We set each emulator parameter to the median value of the training set, and vary each parameter one at a time across the entire prior volume (note that this will result in some unphysical combinations of certain parameters that are strongly correlated). We have not yet varied the slope ($n_p$) of the linear power spectrum in our simulation suites, so do not include it here as an emulated parameter.

## B Redshift evolution of nuisance parameters

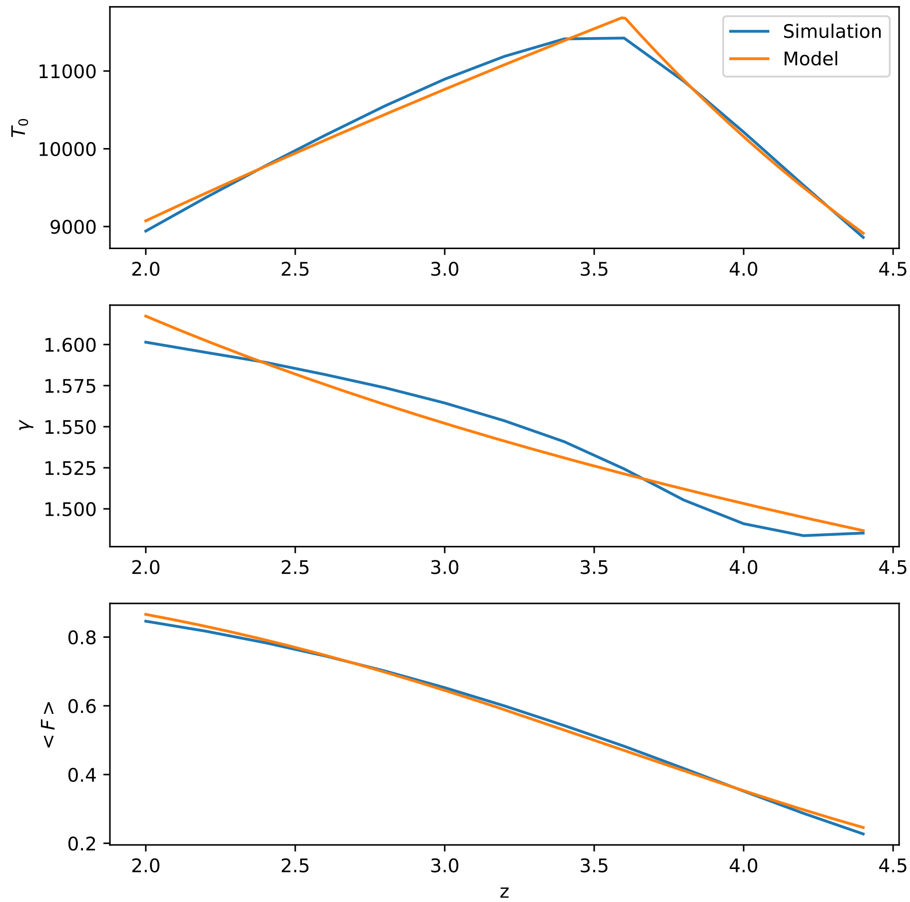


**Figure 3**. The redshift evolution of three of the emulator nuisance parameters for a randomly chosen simulation. In blue we show the measured redshift evolution in the simulation, in orange we show the best fit model using the likelihood models described in section 2.2. The top panel shows the gas temperature at mean density ($T_0[K]$) as a function of redshift, with the spike at $z = 3.6$ caused by He II reionisation. This quantity is converted to the emulator parameter $\sigma_T$.

## C Extensions and unit conversions

Traditionally Ly$\alpha$ forest data is presented in velocity units (km/s), whereas in simulations quantities are measured in terms of Mpc. The conversion factor between these units is $H(z)/(1+z)$, and therefore requires a cosmological model. There are a few possible approaches to deal with this conversion. One is to convert all relevant quantities into km/s at the point of postprocessing the simulation using the $H(z)$ for that particular cosmology, and then work in velocity units in both the emulator and likelihood parameter spaces. A similar approach was taken in [1], however the difficulty here is that the P1D measured at different redshifts and in different simulations will have different $k$ bins in velocity units. This creates an additional layer of complication in the training of an emulator to interpolate the P1D between different simulation points. Instead we measure all emulator quantities in units of Mpc using a fixed simulation box size, meaning that every P1D model has exactly the same $k$ bins. When the emulator calls are made for a likelihood evaluation, we adopt a fiducial cosmology to perform the conversion into velocity units which can be compared with the data. For the final analysis it will be necessary to introduce some parameters to describe some deviation of $H(z)$ and the growth rate from the fiducial cosmology. While the universe is close to Einstein-de-Sitter in the redshfit range $5 > z > 2$, it is not a perfect description, and there is a small variation in the growth rate depending on cosmology (on the order of 2% over the range $0.4 > \Omega_m > 0.2$ at $z = 2$, which is where there is the highest sensitivity). This change in the growth rate affects the Ly$\alpha$ forest through altering the velocities along the lines of sight (which will slightly modify the P1D), and changes the evolution of $\Delta_p$ as a function of redshift.

We could also experiment with looking at the effect of a running of the spectral index, $\alpha$, by adding a relevant parameter to each of the paramter spaces and running simulations with varying $\alpha$.