Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **493,** 4256–4275 (2020)
Advance Access publication 2020 February 20

doi:10.1093/mnras/staa505

Reionization history constraints from neural network based predictions of high-redshift quasar continua

Dominika Ďurovčíková [1]★ Harley Katz,[2]† Sarah E. I. Bosman,[3] Frederick B. Davies [4] Julien Devriendt[2] and Adrianne Slyz[2]

[1]*New College, University of Oxford, Holywell Street, Oxford OX1 3BN, UK*
[2]*Sub-department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*
[3]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[4]*Lawrence Berkeley National Laboratory, Berkeley, CA 94720-8139, USA*

Accepted 2020 February 17. Received 2020 February 17; in original form 2019 November 6

ABSTRACT

Observations of the early Universe suggest that reionization was complete by $z \sim 6$, however, the exact history of this process is still unknown. One method for measuring the evolution of the neutral fraction throughout this epoch is via observing the Lyα damping wings of high-redshift quasars. In order to constrain the neutral fraction from quasar observations, one needs an accurate model of the quasar spectrum around Lyα, after the spectrum has been processed by its host galaxy but before it is altered by absorption and damping in the intervening intergalactic medium (IGM). In this paper, we present a novel machine learning approach, using artificial neural networks, to reconstruct quasar continua around Lyα. Our Quasar Spectra from Artificial Neural Network based predictive Regression Algorithm (QSANNDRA) improves the error in this reconstruction compared to the state-of-the-art principal component analysis (PCA) based model in the literature by 14.2 per cent on average, and provides an improvement of 6.1 per cent on average when compared to an extension thereof. In comparison with the extended PCA model, QSANNDRA further achieves an improvement of 22.1 per cent and 16.8 per cent when evaluated on low-redshift quasars most similar to the two high-redshift quasars under consideration, ULAS J1120+0641 at $z = 7.0851$ and ULAS J1342+0928 at $z = 7.5413$, respectively. Using our more accurate reconstructions of these two $z > 7$ quasars, we estimate the neutral fraction of the IGM using a homogeneous reionization model and find $\bar{x}_{\rm H\,I} = 0.25^{+0.05}_{-0.05}$ at $z = 7.0851$ and $\bar{x}_{\rm H\,I} = 0.60^{+0.11}_{-0.11}$ at $z = 7.5413$. Our results are consistent with the literature and favour a rapid end to reionization.

Key words: intergalactic medium – quasars: emission lines – quasars: general – dark ages, reionization, first stars.

1 INTRODUCTION

The Epoch of Reionization marked a phase transition in the high-redshift Universe during which neutral hydrogen in the intergalactic medium (IGM) became ionized. The history of reionization and sources responsible are two of the major puzzles of modern cosmology. Recent measurements of the cosmic microwave background (CMB) suggest a substantially neutral IGM at $z \gtrsim 7.5$ (Planck Collaboration VI 2018).

Quasars constitute one of the most powerful probes of the IGM at high redshifts due to their extremely luminous and non-transient nature (for a full review, refer to Mortlock 2016). The spectra of these distant objects exhibit a damped Lyα emission profile followed by extensive blueward absorption known as the Gunn–Peterson trough (Gunn & Peterson 1965) arising due to damped and resonant absorption by intervening neutral hydrogen. Both of these features have long been recognized as a useful measure of neutral hydrogen density in the intervening gas; however, Lyα absorption saturates at relatively small neutral fractions making the Gunn–Peterson trough suitable for probing the tail-end of reionization only (Fan et al. 2006). In contrast, studies of the Lyα damping wing (e.g. Bolton et al. 2011; Keating et al. 2015; Davies et al. 2018b) provide a wealth of information on the state of the IGM during reionization.

The prerequisite to extracting this information from the damping wings of high-redshift quasars is the knowledge of their intrinsic

★ E-mail: dominika.durovcikova@gmail.com
† Visitor.

spectra. For the purpose of this work, we define the term *intrinsic spectrum* as the quasar spectrum after it has been processed by its host galaxy but before it has been affected by the intervening IGM. If one has a good model for the intrinsic spectrum of a quasar, one can measure the amount of damping needed to transform this intrinsic spectrum to that observed and thus probe the neutral faction in the vicinity of the quasar.

Fortunately, low-redshift quasars are relatively unaffected by IGM absorption. More than several hundred thousand low-redshift quasars have been observed by the Sloan Digital Sky Survey (SDSS) (York et al. 2000; Eisenstein et al. 2011; Dawson et al. 2013; Blanton et al. 2017; Abolfathi et al. 2018). Strong correlations among the various spectral features in low-redshift quasar spectra have been shown to exist (Boroson & Green 1992; Francis et al. 1992; Yip et al. 2004; Suzuki 2006; Shang et al. 2007). Notably, Eilers et al. (2017) used these correlations to study reionization through principal component analysis (PCA) based proximity zone modelling of $z \sim 6$ quasars (in Eilers, Davies & Hennawi 2018; Eilers et al. 2019). Because the portion of the quasar spectra that is significantly redward of Lyα remains relatively unaffected by the intervening neutral IGM, another approach is to develop a model that relates the region of the spectrum redward to Lyα to that which is blueward. This way, the intrinsic spectrum of a quasar can be reconstructed at high redshift.

Mortlock et al. (2011) and Bañados et al. (2018) used composite spectra of the most similar low-redshift SDSS quasars to reconstruct the intrinsic spectra of two $z > 7$ QSOs. Greig et al. (2017a), Greig, Mesinger & Bañados (2019) constructed a covariance matrix to capture the relationships between the Lyα, Si IV+O IV], C IV and C III] emission lines and used these features to reconstruct the intrinsic spectra of the same QSOs. Davies et al. (2018a) used a PCA technique to extract the mapping from the spectral features redward of Lyα to the blueward features, and once again applied this to the two $z > 7$ QSOs. However, each of these techniques yields different predictions of the shape of the intrinsic spectra, leading to uncertainties on the high-redshift neutral fraction.

The idea of finding relationships between the intrinsic red side and blue side of QSO spectra is well suited to machine learning. Much of the physics that governs this relationship is extremely complicated and not well categorized. For this reason, the mapping between the two regions of a QSO spectrum is non-trivial. However, the low-redshift SDSS QSO data base provides an ideal tool to empirically determine this relationship without making any assumptions on the linearity or the physics involved. In this paper, we present a novel approach to high-redshift spectra reconstruction termed Quasar Spectra from Artificial Neural Network based predictive Regression Algorithm (QSANNDRA). More specifically, we have implemented an ensemble learning technique by combining 100 artificial neural networks (NNs) into a committee to extract the correlations between the regions of a QSO spectrum redward and blueward of Lyα and thus predict the intrinsic spectrum in a $\sim$100 Å long window around the Lyα peak. Due to the entirely empirical nature of the correlations found in quasar spectra, models like these are promising for extracting the complicated correlations. More detailed analysis of the features of the model may also reveal some of the underlying physics of the systems.

We present this work as follows. Section 2 explains how we clean our training data, select an architecture for our NNs, and train our model. In Section 3, we apply our model to two of the highest redshift QSOs known to date, in particular to ULAS J1120+0641 at $z = 7.0851$ (Mortlock et al. 2011; Venemans et al. 2017a), and ULAS J1342+0928 at $z = 7.5413$ (Bañados et al. 2018), and we use the predictions for these two quasars to constrain the neutral fraction by modelling the damping wing profile. Section 4 offers a summary and our conclusions.

2 METHODS

To be able to predict the intrinsic spectrum of a high-redshift quasar with strong absorption blueward of Lyα, we need to train an algorithm on data for which both the red-side and the blue-side spectra[1] are observed with minimum or no absorption. Fortunately, the SDSS (York et al. 2000; Eisenstein et al. 2011; Dawson et al. 2013; Blanton et al. 2017; Abolfathi et al. 2018) has collected data of several hundred thousand low-redshift quasars that are suitable for this purpose. This section explains how we select and clean the low-redshift SDSS data to build our model, while measuring its applicability to high-redshift quasars.

2.1 Data cleaning and training set compilation

In this section, we explain the data cleaning procedure and define the criteria used to select the low-redshift data, thus compiling the preliminary training set for our model.

All low-redshift quasar spectra come from the Extended Baryon Oscillation Spectroscopic Survey (Dawson et al. 2016) of the SDSS-IV (Blanton et al. 2017; Abolfathi et al. 2018) and its earlier phases (York et al. 2000; Eisenstein et al. 2011; Dawson et al. 2013). The primary training data selection was performed based on the 14th data release version of the SDSS Quasar Catalog (Pâris et al. 2018) using criteria that were mostly inspired by Davies et al. (2018b).

In order to define the training set, we first identified all available quasars with `Z_PIPE` redshifts from 2.09 to 2.51. This redshift range enables us to capture the entire Lyα peak as well as other significant features of the spectra, such as the C IV and the Mg II emission lines. We rejected all QSOs with highly uncertain redshifts (`ZWARNING`$\neq$0) and all broad-absorption-line quasars (BALs, `BI_CIV`$\neq$0). These cuts reduced the data set to 101 739 QSOs. We then performed a signal-to-noise ratio cut `SN_MEDIAN_ALL`$>$7.0 which significantly reduced the number of QSOs to 19,054. In Appendix A, we explore the effect that varying the S/N threshold value has on the overall performance of our model.

For each spectrum, we masked out all sky lines listed in table 30 of Stoughton et al. (2002) as well as all pixels that were flagged as highly uncertain. All remaining spectra were subsequently smoothed. A detailed account as well as a visual demonstration of the smoothing procedure is provided in Appendix B, and hence we only provide an outline below for the sake of brevity. We first computed a running median with a bin size of 50 data points to capture the main continuum and emission features in the spectrum. We then performed a peak-finding procedure using the SCIPY PYTHON library (Jones, Oliphant & Peterson 2001) above the aforementioned running median border and interpolated the peaks to construct an upper envelope of the spectrum. This envelope was then subtracted from the spectrum. We then applied the RANSAC regressor algorithm (Fischler & Bolles 1981) from the SCIKIT-LEARN PYTHON package (Pedregosa et al. 2012) on the residuals, thus rejecting most absorption features in the spectrum. The data points that were flagged as inliers by RANSAC were interpolated and

[1]For our purposes, *red-side* denotes wavelengths longer than 1290 Å, while *blue-side* denotes wavelengths shorter than 1290 Å.

smoothed by computing a running median with a bin size of 20, thus creating the final smooth flux fit of each spectrum.

The smoothed spectra were used to perform the final set of cuts. First, the observed wavelengths, $\lambda_{\rm obs}$, were calibrated to rest wavelengths, $\lambda_{\rm rest}$, according to

$$\lambda_{\rm rest} = \frac{\lambda_{\rm obs}}{1+z}, \tag{1}$$

where z is the object's SDSS redshift coming from the broad UV emission lines. In Appendix C, we explore the systematic errors potentially associated with this redshift calibration. Next, we normalized the spectra such that all fluxes at 1290 Å were equal to unity, and then rejected all quasars whose fitted fluxes fell below 0.5 blueward of 1280 Å or below 0.1 redward of 1280 Å. This was done in order to reject quasars with strong associated absorption or poor signal-to-noise ratio redward of Lyα, respectively. It should also be noted that this normalization also removes the sensitivity to the Baldwin effect (Baldwin 1977) as well as the correlation between the quasar brightness and emission line shifts (Shang et al. 2003; Richards et al. 2011). The significance of this step is discussed in Section 4. The remaining data set consists of 17 007 quasars, whose fluxes were interpolated at 3862 different wavelengths between 1191.5 and 2900.0 Å (spaced uniformly in log space).

2.2 Refining the training set with random forests

Visual inspection of the data revealed that our training set still contained a few spectra with strong absorption features blueward of 1290 Å. This section describes a random forest (RF) based procedure we used to further reject these quasars from the training set.

RF regression is an ensemble learning algorithm, which combines multiple decision trees to form a statistical prediction of the output value, or the blue-side flux in our case. Within the RF, each decision tree makes a flux estimation after a series of queries on the red-side spectral properties. The overall predicted flux is then determined by taking the average of predicted flux values by all the trees in the forest.

The power of RFs results from the way the trees are grown in the training phase. Growing each individual tree is done by a random selection of spectral features and training spectra (i.e. bagging or bootstrapping) based on which the tree learns to make the prediction. The training is performed by feeding example, or training, spectral data into the forest, where each tree gradually learns to map the input (red-side) values to the output (blue-side) ones by comparing its own prediction to what the output is supposed to be.

The motivation behind choosing a RF regressor (Breiman 2001), as implemented in the SCIKIT-LEARN PYTHON package (Pedregosa et al. 2012), is that RFs are rather easily trained and are useful in confirming strong correlations between the red-side and the blue-side spectral features. Because of these correlations, it is likely that the RF would be unable to predict uncorrelated absorption features, which makes its predictions a suitable tool to detect the remaining quasars with strong absorption in the training set. In other words, the RF will have a large prediction error on the few remaining QSOs in our data set that have strong absorption features around Lyα and thus this high prediction error is indicative of an outlier in our data set. Furthermore, because each tree learns the red-side to blue-side mapping based on a different set of criteria and training spectra, we can prevent overfitting even when using decision trees that are arbitrarily deep. The downsides are that training large RFs is extremely memory-intensive and that they may not generalize as well as other machine learning algorithms (such as a NN).

To train a RF, we split the training set into train and test subsets (80:20 ratio). We standardized the fluxes to have a mean of 0 and a variance of 1 for each wavelength across all objects, upon which we performed PCA using the SCIKIT-LEARN PYTHON package (Pedregosa et al. 2012) to reduce the feature space and complexity of the RF. Standardization was performed in order to make the different quasars as well as the different features in each spectrum comparable, which improves the PCA search for vectors of maximum variance. We chose the number of principal components such that they capture 99 per cent of variance in the data both on the red side and the blue side (we will henceforth refer to this number as the *explained variance ratio*). After this transformation, the data were fed into a RF with 100 decision trees.

Inspection of the preliminary RF predictions revealed that, as expected, the RF predicted large errors for the QSOs in the data set that exhibited strong absorption features. We took advantage of this feature to remove most of the remaining quasars with strong absorption from our training set as follows. We defined the relative prediction error, ϵ, to be the relative absolute error of the prediction against the smoothed flux value at a particular wavelength, or

$$\epsilon = \frac{|\, F_{\rm pred} - F_{\rm smooth} \,|}{F_{\rm smooth}}, \tag{2}$$

where $F_{\rm pred}$ is the flux predicted by the RF and $F_{\rm smooth}$ is the smoothed flux. Based on the error for each predicted data point in the test set, we rejected all data points whose relative prediction error was greater than $\bar{\epsilon} + 3\sigma_\epsilon$, where $\bar{\epsilon}$ is the mean error and σ_ϵ is the standard deviation of the error across all objects for a particular wavelength. This procedure was repeated 10 times, each time training the RF on nine subsets of the whole training set (henceforth termed *folds*) and rejecting data points in the 10th one. This way, 1.1 per cent of all data points on the blue side were rejected in the whole training set, altogether rejecting 3304 objects. This left us with a cleaner training set of 13 703 spectra. Fig. 1 displays two typical examples of spectra that were rejected in this process.

2.3 Construction of QSANNDRA

This section outlines the implementation of a feed-forward NN on our training set and further describes the construction and training of our predictive model called QSANNDRA.

We implemented an artificial NN in order to better capture the correlations between the different spectral features of QSOs. NNs are stacks of interconnected layers of computational units called neurons, where each neuron is assigned a set of weights, w_i, and a bias, b. It multiplies all its inputs, x_i, by the corresponding weights and adds the bias before passing the result on to the following neurons in subsequent layer. Each layer is then assigned an activation function f, so that the output of a neuron in that layer, y, becomes

$$y = f\left(b + \sum_i w_i x_i\right) = f(b + \boldsymbol{w}^{\rm T}\boldsymbol{x}). \tag{3}$$

The power of even the simplest NNs is that they allow for modelling of non-linear relationships between the input and the output. This is due to the fact that the information from the different neurons is combined in a non-linear fashion as it propagates through the layers of the network. In an untrained network, the weights and biases on each neuron are typically initiated randomly. Then, by iteratively

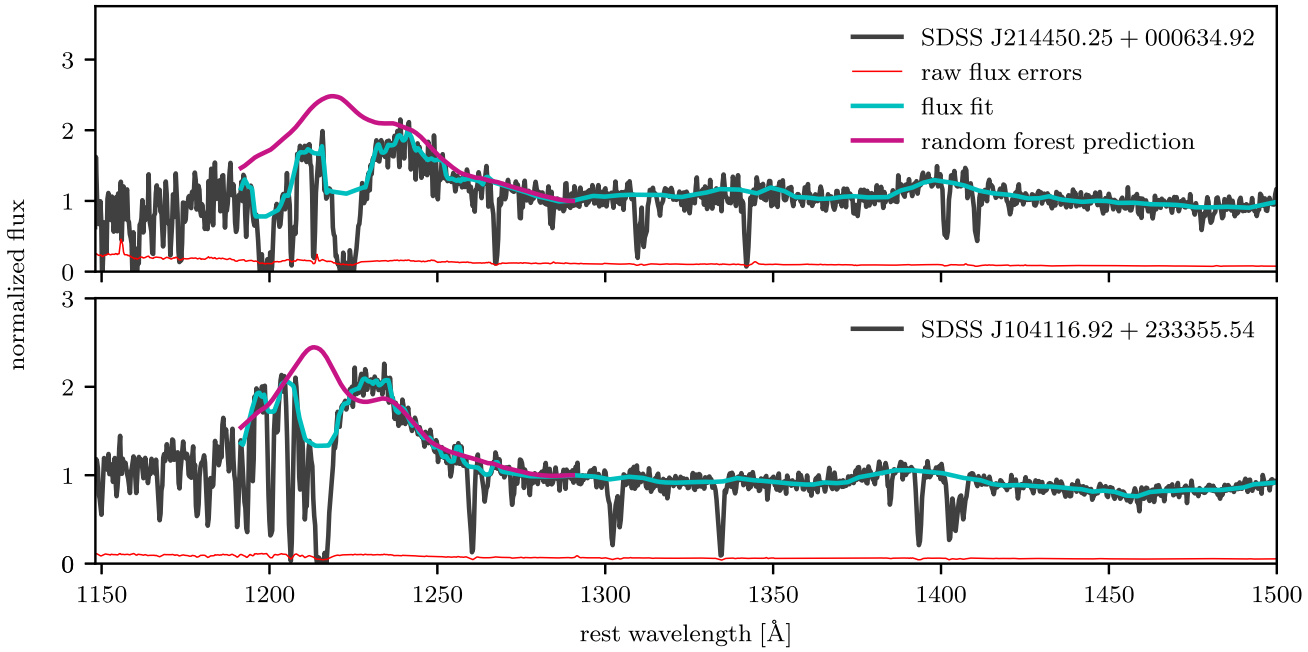

Figure 1. Two examples of spectra with strong absorption features blueward of 1290 Å that were rejected based on preliminary RF predictions. The raw smoothed data for each quasar are shown in grey, the flux fit is shown in cyan and our blue-side predictions are shown in magenta. We also show the flux errors on the raw data as a thin red curve. Note that the *y*-axis is normalized with respect to each quasar's fitted flux at 1290 Å.

passing training inputs (i.e. red-side spectra) through the network, comparing its outputs (i.e. blue-side spectra) to what the training outputs should be and updating the weights and biases along the gradient of the loss function[2] with respect to that weight, learning of the ANN is achieved. More detailed information on NNs can be found in Géron (2017).

We implemented a fully connected feed-forward NN using the KERAS PYTHON package (Chollet et al. 2015), which means that each neuron was connected to all neurons in the preceding and following layer, and that information was propagated in only one direction from the input to the output. Since we operate with 63 and 36 principal components on the red side and the blue side, respectively, the number of neurons in the input and output layers was fixed at 63 and 36. To further define its architecture and hyperparameters, i.e. number of layers, number of neurons in each layer, activation function, number of training epochs[3] and the batch size,[4] we performed an extensive grid search over 100 different networks and training settings. For each combination, 10-fold cross-validation (see Fig. 2) was performed to ensure generalizability, i.e. a comparable performance when applied to previously unseen data. For cross-validation, the training subset was further divided to 10 folds, and the NN training was repeated 10 times, each time training on 9 folds and computing the mean-absolute-error (i.e. the loss function) of the prediction on the 10th fold. The mean error and its standard deviation was reported for each NN, and these scores were then compared to choose the NN with the best performance. To further fine-tune the training parameters, i.e. number of epochs and

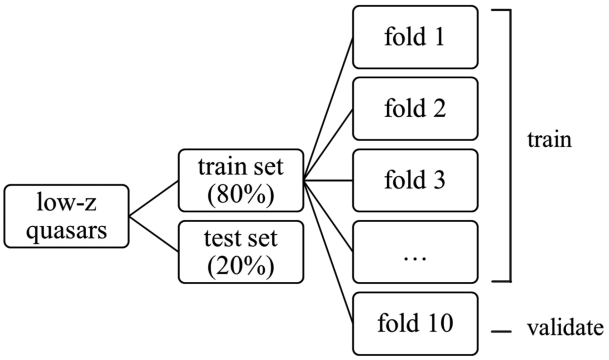

Figure 2. Infographic showing how the low-redshift data were divided for 10-fold cross-validation. The whole training set was divided into a train and a test set (80 per cent and 20 per cent of the quasars, respectively), and the train set was further subdivided into 10 folds. During cross-validation, the NN was trained 10 times, each time training on 9 folds and validating its performance on the remaining fold to assess generalizability.

the batch size, a further, smaller grid search over these parameters was performed for the best-performing NN architecture.

The NN with the best performance was found to have the following architecture: 63-40-40-36[5] with the 'elu' activation function,[6] and was trained for 80 epochs using a batch size of 500 QSOs. When applied to the test set, the mean relative prediction error was 5.7 per cent per data point. Fig. 3 displays how $\bar{\epsilon}$ (solid curves) and

[2]The loss function defines the discrepancy between the predicted flux value and the real flux value.

[3]One epoch is defined as passing the full training data set through the network once.

[4]The batch size is the number of training examples that are passed through the network before the weights get updated.

[5]Each number stands for the number of neurons in the corresponding layer.

[6]The 'elu' function is defined as

$$f(x) = \begin{cases} x & x > 0 \\ \alpha(e^x - 1) & x \leq 0 \end{cases},$$

where -α defines the horizontal asymptote at negative infinity.

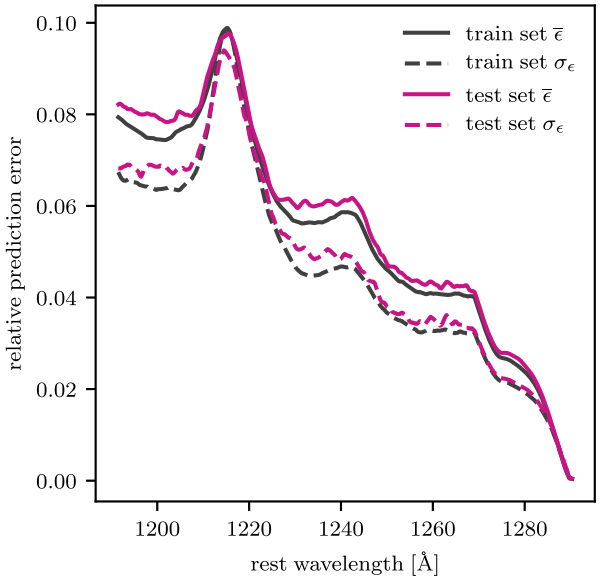

Figure 3. Mean (solid curve) and standard deviation (dashed curve) of the relative prediction error as defined in equation (2) for the NN blue-side predictions as a function of wavelength. The set of grey curves shows the NN's performance on the train set, while the set of magenta curves shows its performance on the test set (i.e. previously unseen data).

σ_ϵ (dashed curves) vary across the predicted wavelength range both for the train set (grey set of curves) and the test set (magenta set of curves). Note how there are distinct features in the error of our model as a function of rest-frame wavelength. In general, the redder the wavelength, the less error we predict for our model. This is simply due to the fact that the closer the wavelength is to the known data, the more reliable the extrapolation is. The bumps in the error are due to the presence of emission lines in the QSO spectra. Most notably, the three bumps that we see are due to Lyα, N V at 1240 Å, and Si II at 1260 Å, from blue to red respectively. The strengths and velocity shifts of these emission lines are more difficult to predict than the underlying continuum and hence the error is enhanced around their wavelengths. Fig. 4 shows two examples of low-redshift quasars drawn from the test set, where we have used our NN and the earlier trained RF to predict the QSO continuum around Lyα. In both examples, our predictive model performs very well.

To improve the performance and increase the robustness of our method, we trained 99 more NNs with the same architecture and hyperparameters. The weights and biases of each of the neurons were initiated using a different random seed. This way, we created a committee of NNs, which falls under ensemble learning techniques (see Dietterich 2000). The idea behind ensemble learning is that if the errors made by individual predictors are uncorrelated, they will cancel out with each other when averaging the predictions from multiple algorithms. Furthermore, it is possible for a NN to get trapped in a local minimum during training and thus never reach the optimal solution. Initiating the weights and biases in each NN differently provides a different starting point for each training, which can result in different networks converging to different local minima. Averaging these locally optimal predictions has the potential to come nearer to the globally optimal prediction and thus decrease the overall prediction uncertainty.

We trained each network separately and evaluated the performance of each of them on the test set of the low-redshift SDSS data. The inverse of the achieved mean relative prediction error $\bar{\epsilon}$ then determined the weight assigned to each of the individual NNs in the ensemble. Subsequently, the weighted predictions were averaged to produce the overall prediction of the committee, whose performance was once again evaluated on the low-redshift test set. This method improved the mean relative prediction error to 5.5 per cent per data point (as compared to 5.7 per cent for the individual NN described previously).

It should be emphasized that the power of this approach lies in the potential of the ensemble outperforming even the best algorithm within the ensemble. In fact, as shown in Fig. 5, the committee achieved a ∼5 per cent improvement on the mean relative prediction error around the Lyα peak as compared to the best-performing NN in our ensemble for both the train and test set predictions. Moreover, we observe that the ensemble achieves a mean relative prediction uncertainty spanning from ∼5 per cent to ∼9 per cent in the wavelength range ∼1210 to ∼1250 Å most relevant to damping wing modelling, with the standard deviation spanning from ∼4 per cent to ∼7.5 per cent. The difference between the performance of the committee on the train and test sets is marginal, which suggests a strong generalizability of our model to new data. The validation stage of the ensemble's performance thus confirms an improved accuracy and robustness in using the committee of networks as opposed to just one individual prediction.

2.4 Applicability of QSANNDRA to $z > 7$ QSOs

Although we have developed a robust algorithm for predicting the intrinsic QSO spectra of low-redshift SDSS QSOs, it is important to determine the applicability of this model to the high-redshift quasars that we aim to use to constrain the neutral gas fraction during the Epoch of Reionization. In particular, we aim to apply our model to the combined VLT/FORS + Gemini/GNIRS spectrum of ULAS J1120+0641 (Mortlock et al. 2011) and the Magellan/FIRE + Gemini/GNIRS spectrum of ULAS J1342+0928 (Bañados et al. 2018). If there is a fundamental difference between the high-redshift quasars and the SDSS QSOs, even though the trained NNs are meant to be generalizable, their predictive power on such different systems deserves to be questioned. For this reason, in this subsection, we provide a method to determine both how similar the two $z > 7$ QSOs are to the SDSS quasars as well as the performance of QSANNDRA on the low-redshift QSOs that are most similar to those at $z > 7$.

To quantify how unusual the two high-redshift quasars are, we trained an autoencoder on the red-side PCA components of the low-redshift SDSS data. An autoencoder is a NN with two components, an encoder and a decoder. The first compresses the input data while the second reconstructs it. It essentially acts as an identity function. Training the autoencoder on the SDSS QSOs causes the network to pick out the spectral features that are most represented in our low-redshift training set, which then form the basis for reconstruction. By measuring the error between the input and the reconstructed output coefficients, one can determine how effective the compression of the data is for all the SDSS quasars, and hence define an error distribution that is characteristic of our low-redshift sample. Subsequently, by running the trained autoencoder on the red-side PCA coefficients of the two high-redshift quasars and measuring the error, we are able to quantify how well represented the red sides of the high-redshift quasars are in our training data set and hence determine the extent of their outlying nature.

We thus trained a 4-layer autoencoder whose input and output layers both had 63 neurons (corresponding to 63 red-side PCA components) and the middle two layers were composed of 30

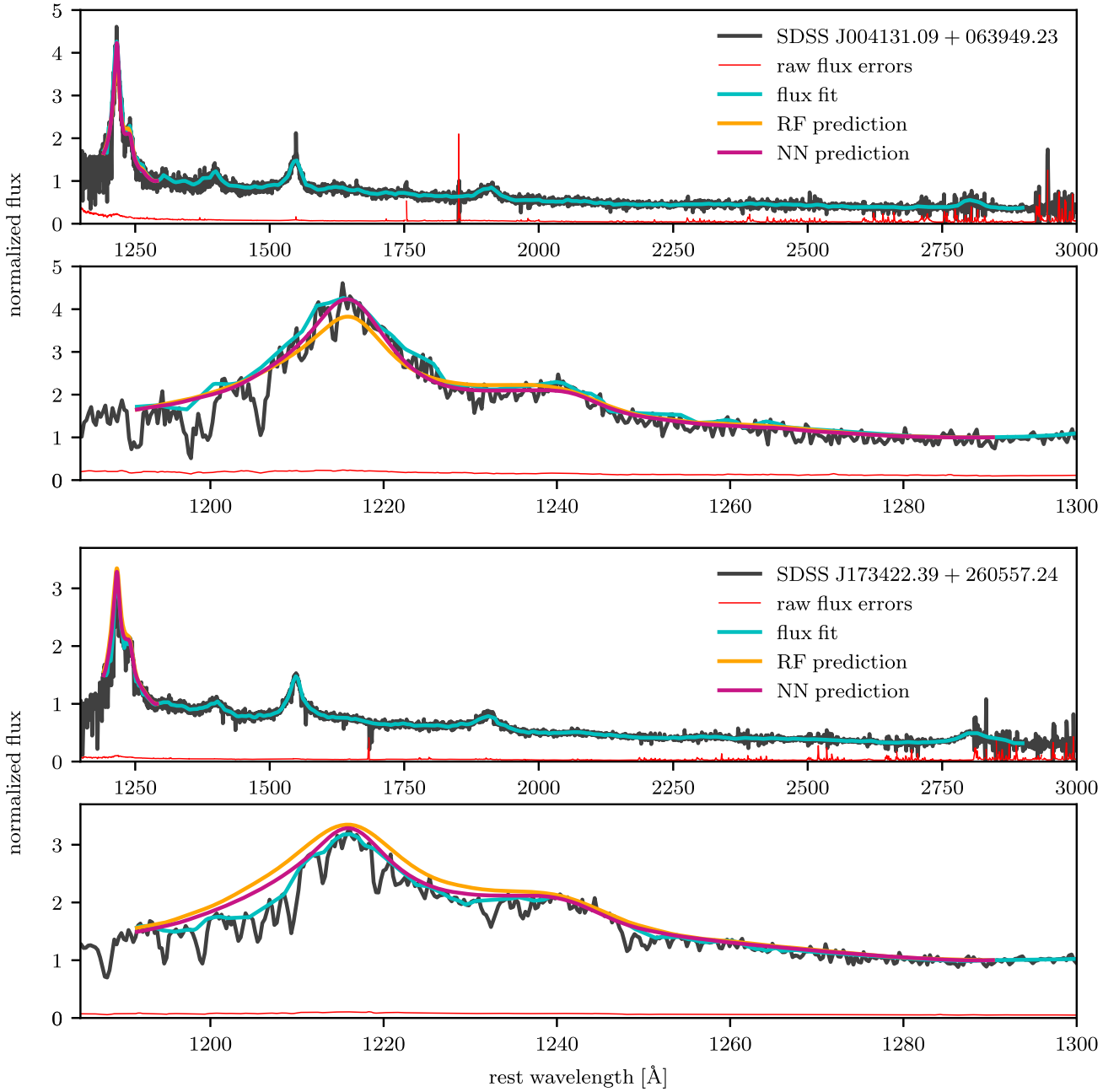

Figure 4. Two examples of low-redshift quasar spectra drawn from the test set comparing the RF (orange) and NN (magenta) predictions. The raw data are depicted in grey, the errors on the raw data are shown as a thin red curve, and the flux fit is shown in cyan. The top panel for each quasar shows the full spectrum, while the bottom panel offers a close-up view of the predictions. Note that the *y*-axis is normalized with respect to each quasar's fitted flux at 1290 Å.

neurons. The activation function on all neurons was set to 'elu'. We trained the autoencoder for 100 epochs in batch sizes of 500 while optimizing for the mean squared error between the input and output coefficients. We then applied the trained autoencoder on the test set of low-redshift quasars. We further calculated the resulting mean squared error across all red-side coefficients for each quasar both in the train set and the test set and composed a probability density function for this error for the low-redshift data set.

Applying the trained autoencoder to ULAS J1120+0641 and ULAS J1342+0928 and calculating their corresponding mean squared errors revealed that they fall on to the 48th percentile and the 99th percentile in our low-redshift distribution (see Fig. 6). While ULAS J1120+0641 seems to be well represented in our low-redshift data set, ULAS J1342+0928 falls on to the tail-end of our distribution. These findings are comparable to those of Davies et al. (2018a), who used a mixture of multivariate Gaussians to determine a percentile of 15 per cent and 1.5 per cent for the two high-redshift QSOs with respect to their low-redshift data set (values equivalent to our percentiles of 85 per cent and 98.5 per cent). While these agree that the $z = 7.5$ QSO is a 2σ outlier, our training set seems to be more representative of the $z = 7.1$ QSO than that of (Davies et al. 2018a). Especially in light of the outlying nature of ULAS

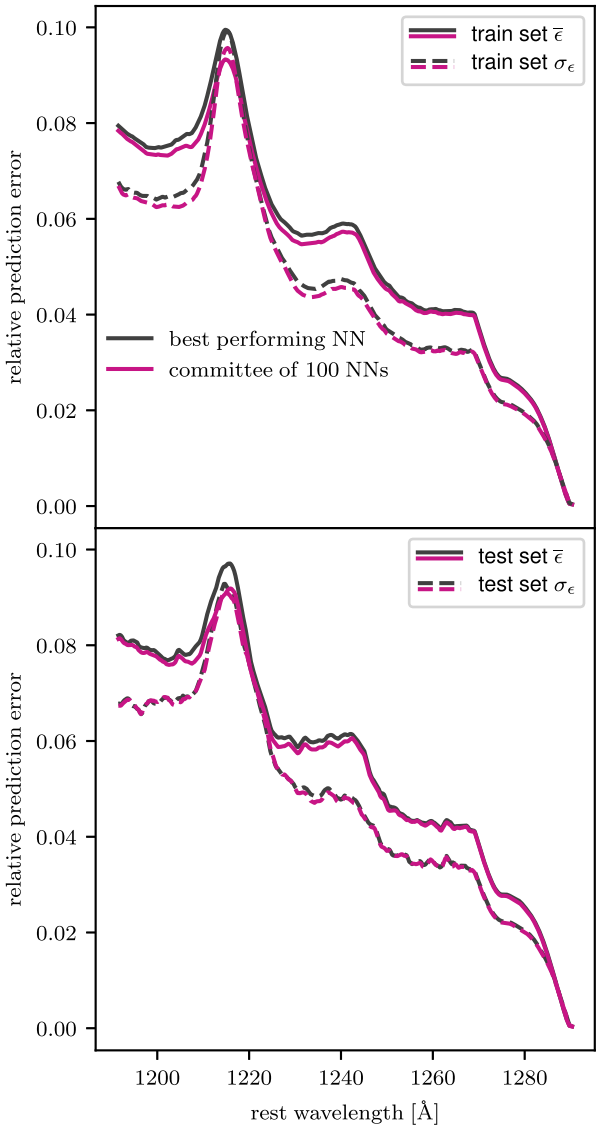

Figure 5. A comparison of the relative prediction error statistics as defined by equation (2) for the committee of 100 networks (magenta) to the best-performing NN (grey) within the committee for the train set (top) and the test set (bottom). The solid curves show the mean relative prediction error, and the dashed curves represent its standard deviation. The ∼5 per cent decrease in mean relative error around the Lyα peak in both train and test sets confirms the improvement in both accuracy and robustness upon implementing the ensemble technique.

J1342+0928, it is crucial to assess the performance of our model on its nearest-neighbour low-redshift quasars.

Motivated by Fig. 6, we further evaluated QSANNDRA's performance on 100 nearest-neighbour QSOs of ULAS J1120+0641 and ULAS J1342+0928 (i.e. the QSOs that are most similar to the high-redshift quasars). We chose the 100 quasars by computing the Euclidean distance between the red-side PCA coefficients of each of the high-redshift quasars and the red-side PCA coefficients of the low-redshift quasars in our test set only. Only searching through the test set of SDSS quasars, which the model did not see during the training stage, guarantees a generalizable performance. The resulting $\bar{\epsilon}$ and σ_ϵ distributions across the blue-side wavelengths are shown in Fig. 7, where the performance on 100 nearest neighbours of ULAS J1120+0641 and ULAS J1342+0928 are shown in orange

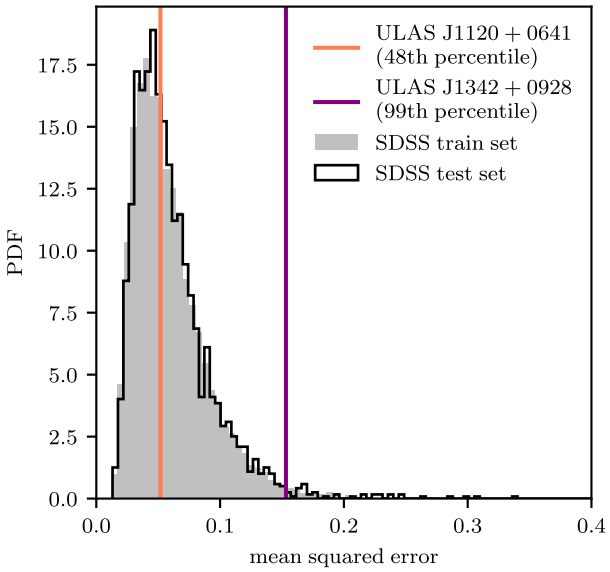

Figure 6. A histogram of autoencoder-produced mean squared error for our low-redshift data set as compared to the mean squared error for ULAS J1120+0641 and ULAS J1342+0928. The distribution of the train set is displayed in grey, the distribution of the test set is displayed in black, while the orange and purple lines indicate the position of ULAS J1120+0641 (48th percentile) and ULAS J1342+0928 (99th percentile), respectively, within that distribution. Note that the percentile here defines how well the spectral features of the two high-redshift quasars are represented in the training sample of low-redshift SDSS quasars (i.e. the larger the percentile, to more outlying the quasar with respect to the low-redshift data set).

and purple, respectively, and the full test set performance is shown in magenta for comparison. QSANNDRA performs better on SDSS quasars that are similar to the high-redshift ones (the full test set $\bar{\epsilon}$ is 5.5 per cent on average as compared to 4.3 per cent and 4.4 per cent on average for the nearest neighbours of ULAS J1120+0641 and ULAS J1342+0928, respectively; the full test set σ_ϵ is 4.8 per cent on average as compared to 3.3 per cent and 3.7 per cent on average for the nearest neighbours of ULAS J1120+0641 and ULAS J1342+0928, respectively) than on the full test set, which hints at a greater reliability of our predictions for these two high-redshift quasars and confirms that QSANNDRA should be able to tackle the outlying spectral features of ULAS J1342+0928.

2.5 QSANNDRA compared to existing models

Before proceeding, it is important to understand how the predictive power of the QSANNDRA model compares to other models available in the literature. To better quantify the performance of QSANNDRA to existing models, we implemented the state-of-the-art PCA-based model from Davies et al. (2018a) and an extension thereof (explained below) on our cleaned training data set to directly compare the results. Note that there are some differences in our implementation compared to the model published by Davies et al. (2018a) which makes it more comparable to the traditional PCA-based techniques (Suzuki et al. 2005; Pâris et al. 2011). Our cleaning and smoothing procedures are slightly different, we do not use nearest-neighbour stacks to compute the PCA basis vectors, and we do not fit for redshifts simultaneously. In contrast to the original paper (Davies et al. 2018a), our training data set relies on a different version of the SDSS data base, and we also split our low-redshift SDSS data set into train and test sets to ensure that we are making

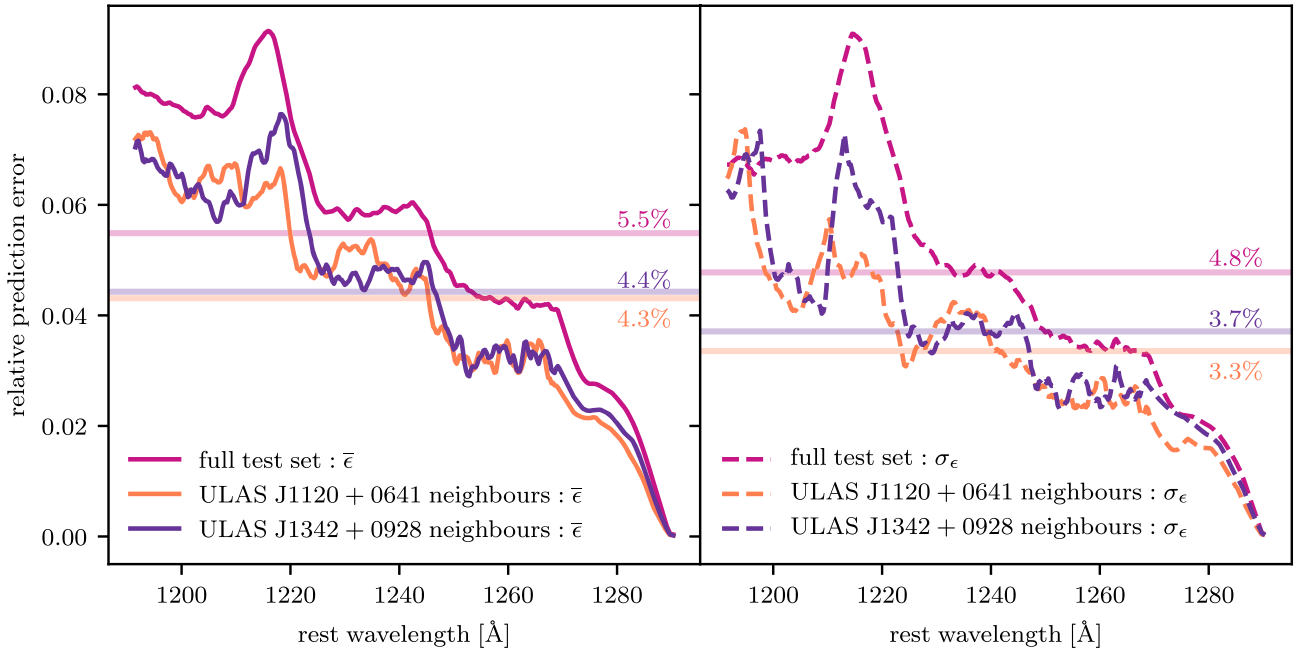

Figure 7. Wavelength-dependent distribution of $\bar{\epsilon}$ (left) and σ_ϵ (right) for QSANNDRA's full test set performance (magenta) as compared to its performance on 100 nearest-neighbour low-redshift quasars from the test set for ULAS J1120+0641 (orange) and ULAS J1342+0928 (purple). Note that even though the $z = 7.5$ QSO has been shown to be on the tail-end of the low-redshift quasar distribution, QSANNDRA actually performs better on low-redshift quasars which are similar to it.

a fair comparison between the models' performances on previously unseen data.

We implemented the model by Davies et al. (2018a) with 10 red-side and 6 blue-side PCA components, and solved for a linear mapping between the red-side and the blue-side coefficients using a least-squares solver. We then computed the blue-side coefficients for all quasars in the test set and $\bar{\epsilon}$ and σ_ϵ for each blue-side wavelength in each case.

As an extension to the original model (Davies et al. 2018a), we adjusted the number of principal components in this model to be the same as in our model to work with an equivalent amount of information (63 on the red side and 36 on the blue side). We will henceforth refer to this model as the extended PCA model. We repeated the same procedure as described above to assess the improvement of QSANNDRA on the state-of-the-art model more fairly. Finally, we also computed $\bar{\epsilon}$ and σ_ϵ for the 100 nearest low-redshift quasar neighbours of both ULAS J1120+0641 and ULAS J1342+0928 according to the same procedure as outlined in 2.4, and compared the results to QSANNDRA.

Fig. 8 displays the ratio of QSANNDRA's relative prediction errors to our calculated relative prediction errors for the original model (Davies et al. 2018a) and the extended PCA model described above. On average across all target wavelengths, we achieve a 14.2 per cent improvement in $\bar{\epsilon}$ (left) and a 11.5 per cent improvement in σ_ϵ (right) on the full test set as compared to Davies et al. (2018a) (shown in grey). When compared to the extended PCA model, we achieve a 6.1 per cent improvement in $\bar{\epsilon}$ (left) and a 4.9 per cent improvement in σ_ϵ (right) on the full test set (shown in magenta) with the least improvement being at the Lyα peak. More interestingly, we can compare the performances of QSANNDRA to the extended PCA method on the QSOs in the SDSS data set that are most similar to the $z > 7$ QSOs that we will use to constrain the high-redshift neutral gas fraction. To do this, we selected the 100 quasars most similar to ULAS J1120+0641 and another 100 most similar to ULAS J1342+0928 based on the red side of the spectra. On these subsets, QSANNDRA's performance further improves with respect to the extended PCA method. In particular, on the 100 quasars most similar to ULAS J1120+0641 (at $z = 7.085$, shown in orange), we achieve a 22.1 per cent improvement in $\bar{\epsilon}$ (left) and a 26.2 per cent improvement in σ_ϵ (right). On the 100 nearest neighbours of ULAS J1342+0928 (at $z = 7.5413$, shown in purple), we achieve a 16.8 per cent improvement in $\bar{\epsilon}$ (left) and a 17.5 per cent improvement in σ_ϵ (right). This experiment indicates that overall, the QSANNDRA is more predictive than the model presented in Davies et al. (2018a) as well as its extension and that this is especially true for the two $z > 7$ QSOs under consideration.

3 APPLICATION TO HIGH-z QUASARS

Given that we are now confident that our model generalizes to high-redshift objects and performs better than other models in the literature, in this section, we apply QSANNDRA to two high-redshift quasars, namely ULAS J1120+0641 at $z = 7.0851$ (Mortlock et al. 2011; Venemans et al. 2017a), and ULAS J1342+0928 at $z = 7.5413$ (Bañados et al. 2018) to reconstruct their blue-side spectra and constrain the neutral gas fraction at their corresponding redshifts.

3.1 Reconstructing high-redshift quasar spectra

In order to apply our algorithm to the high-redshift QSOs, we need to ensure a good fit of the red-side continuum and emission features. Notably, the spectrum of ULAS J1120+0641 contains a region of poor S/N between ∼1660 and ∼1800 Å, and between ∼2200 and ∼2450 Å, and the spectrum of ULAS J1342+0928 has missing data between ∼1570 and ∼1700 Å, and between ∼2100 and ∼2230 Å.

To model these parts of the respective spectra as accurately as possible, we took advantage of the correlations between the various

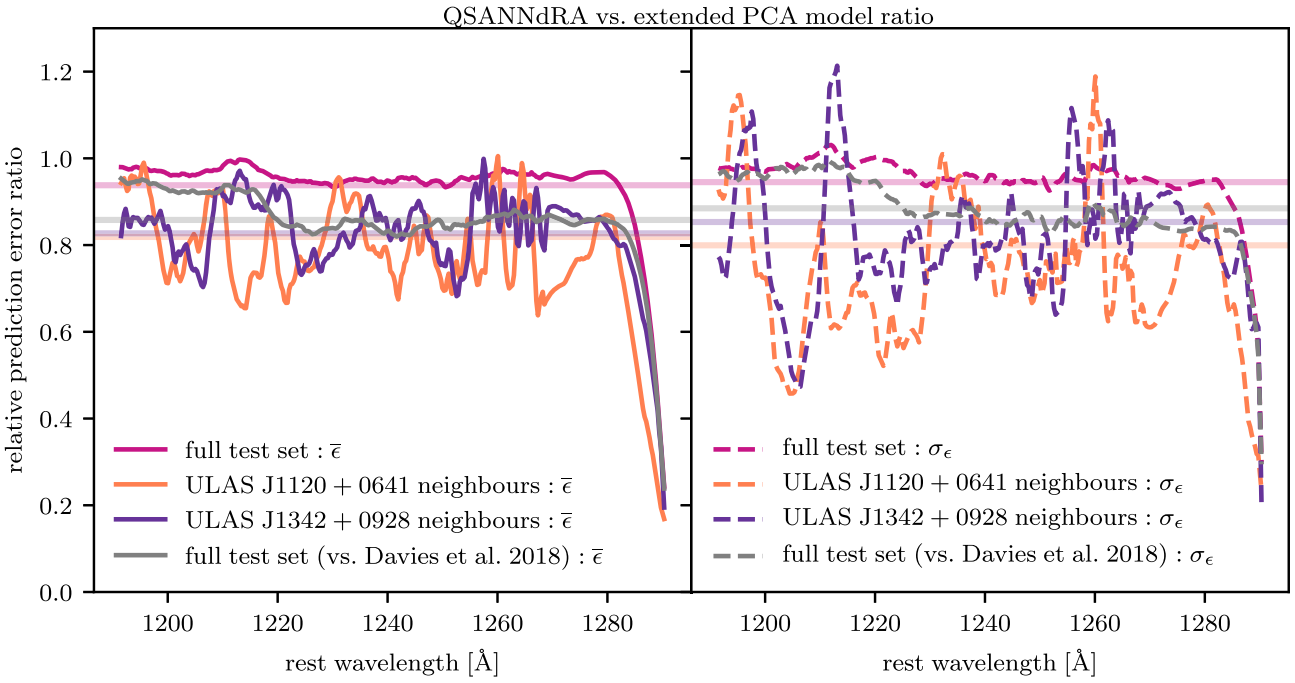

Figure 8. Wavelength-dependent distribution of the ratio of QSANNdRA's performance to the original PCA model's performance reported by Davies et al. (2018a) as well as the extended PCA model's performance in terms of $\bar{\epsilon}$ (left) and σ_ϵ (right). The full test set performance improvement as compared to the published model (Davies et al. 2018a) is shown in grey. When compared to the extended PCA model, the full test set performance ratio (magenta) is compared to the performance ratios on the 100 nearest low-redshift quasar neighbours of ULAS J1120+0641 (orange) and ULAS J1342+0928 (purple) from the test set.

features in the spectra again. We trained two very simple NNs for the two cases (with an architecture of 55-20-11 neurons and training in batches of 800 for 400 epochs with the 'elu' activation function), which learnt to predict the poor or missing data based on the remaining parts of the red-side spectra using the training set of low-redshift SDSS QSOs described in Sections 2.1 and 2.2. The resultant predictions had both the mean error and its standard deviation below 2.5 per cent for all target wavelengths in both cases and displayed a strong generalizability to previously unseen data.

With the fully reconstructed red-side spectrum, we performed the same fitting procedure as outlined in Section 2.1 and finally applied the trained committee of networks from Section 2.3 to reconstruct the blue-side spectrum for each $z > 7$ QSO.

In Fig. 9, we show our reconstruction of the continuum spectra of the two $z > 7$ QSOs, ULAS J1220+0641 and ULAS J1342+0928. This figure shows the observed quasar spectra in grey along with the uncertainties in each observed data point in red. The cyan curve shows our fit of the red-side spectra. We then show the individual predictions from the 100 NNs in the committee as light magenta curves and emphasize the resultant weighted prediction of the committee as the thick magenta curve. We also include the full test set $\bar{\epsilon}$ uncertainty bounds on our final predictions (see Fig. 7) as the dotted magenta curves. As expected, the cyan curve that represents the red-side fits the observed data extremely well, even in the regimes where there is low S/N or missing data. For the $z = 7.1$ QSO, ULAS J1120+0641, the reconstructed continuum has a weak peak near Lyα compared to the observed spectra indicating that limited absorption is occurring at the Lyα peak. In contrast, for the $z = 7.5$ QSO, ULAS J1342+0928, the reconstructed continuum predicts a significantly stronger Lyα peak than what is observed. Similarly, this QSO also sees more enhanced emission at the location of the NV doublet emission line at a rest-frame wavelength of ∼1240 Å compared to the slightly lower redshift object. Despite the fact that the SiII emission line at ∼1262 Å is visible in the error plots for our predictive model, in neither high-redshift QSO do we predict this emission line in reconstruction.

It is also important to note that all NNs in our model make a prediction that the $z = 7.5$ QSO should have had a strong NV line. The fact that this discrepancy between our prediction and the observed flux is much larger than the ∼ 5 per cent uncertainty predicted in Fig. 7 weakens the reliability of the continuum prediction in this case. This discrepancy also impacts the interpretation of the neutral fraction constraint presented in the next section, as more neutral gas is necessary to reconstruct the observed spectrum from our prediction.

In Fig. 10, we compare our reconstructed QSO spectra for the two $z > 7$ QSOs with other predictions from the literature. In the left-hand panel of Fig. 10, we show how our prediction for the blue-side spectrum of ULAS J1120+0641 (magenta) compares to the predictions based on the SDSS composite (orange) (Mortlock et al. 2011), the covariance matrix approach (green) (Greig et al. 2017a), and the PCA method (blue) (Davies et al. 2018a). Interestingly, each of the different reconstruction methods gives a different prediction for the intrinsic spectrum. Our method yields a prediction that is more similar to that of the SDSS composite (Mortlock et al. 2011) as well as that predicted in Davies et al. (2018a) compared to that predicted in Greig et al. (2017a). As noted earlier, we have predictions from 100 individual NNs that go into our ensemble. None of these 100 NNs predict a Lyα peak that is nearly as strong as that from Greig et al. (2017a). In contrast, some of the individual NNs do predict Lyα as strong as that seen in the SDSS composite and the PCA method. Our model for this QSO can in general be categorized as having the weakest Lyα and NV emission, the

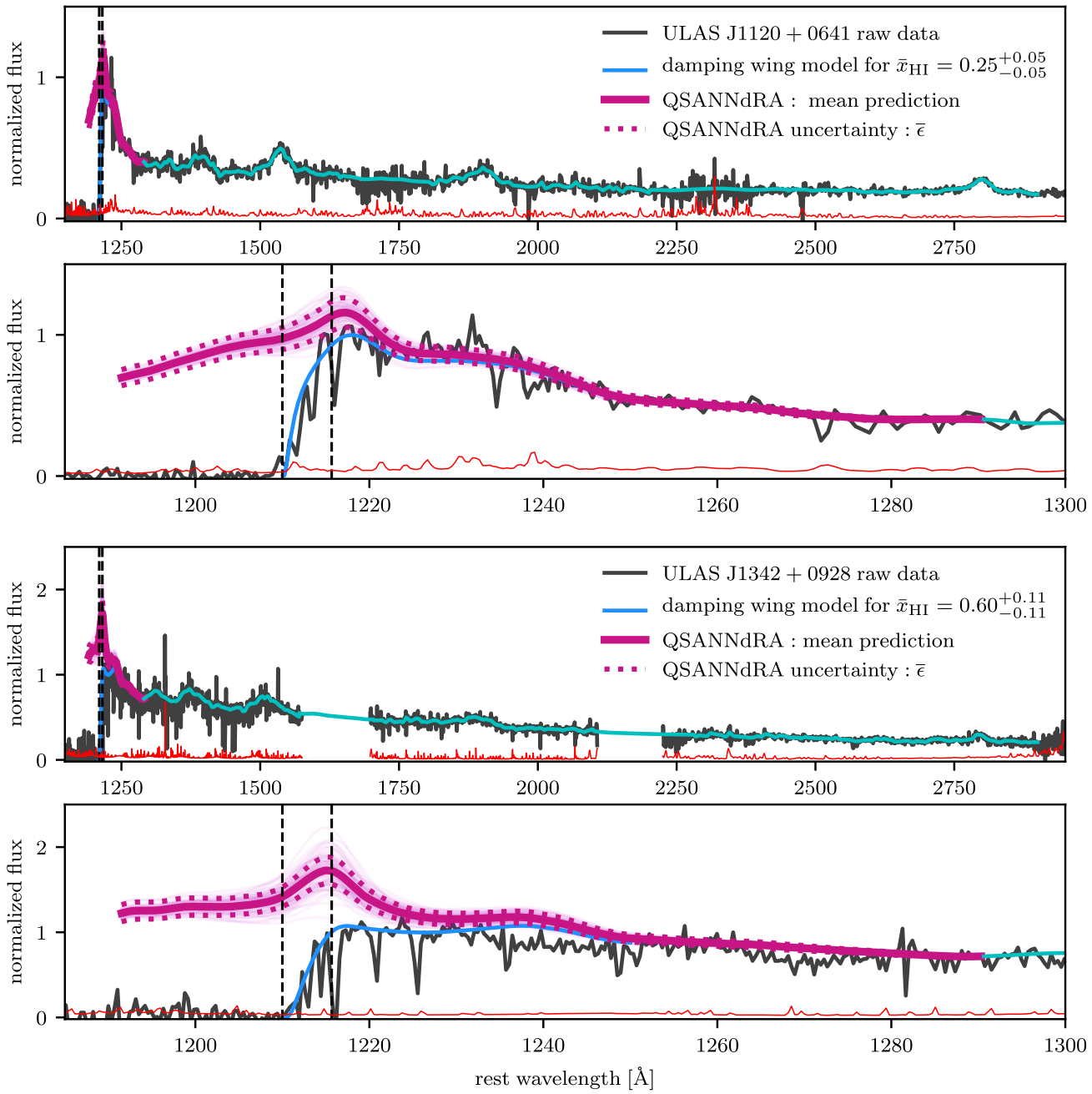

Figure 9. The reconstructed spectrum of ULAS J1120+0641 (top two panels) and ULAS J1342+0928 (bottom two panels) with a model of the damping wing. The bottom panel for each quasar shows a close-up view of the Lyα region. The raw data points and their uncertainties are shown in grey and red, respectively. The cyan curve represents our fit of the red-side spectrum. For ULAS J1220+0641, the flux in the poor S/N regions between ∼1660 and ∼1800 Å, and between ∼2200 and ∼2450Å was reconstructed based on the low-redshift QSO spectra. For ULAS J1342+0928, the flux in the regions of missing data between ∼1570 and ∼1700Å, and between ∼2100 and ∼2230Å were also reconstructed based on the low-redshift QSO spectra. The thin light magenta lines show the individual predictions from the 100 NNs with the committee, while the thick magenta line shows the weighted average of these predictions at each wavelength. We also show the full test set $\bar{\epsilon}$ bounds on our predictions from Fig. 7 as the dotted magenta curves. The damping wing model is shown in blue and corresponds to $\bar{x}_{\rm HI} = 0.25$ for ULAS J1220+0641 and $\bar{x}_{\rm HI} = 0.60$ for ULAS J1342+0928, which was calculated as the weighted average of optimal neutral fractions corresponding to the individual predictions of the 100 networks within the committee. The region between the vertical dashed black lines represents the QSO proximity zone as defined in Section 3.2.

latter being more consistent with the SDSS composite than that of the PCA method. Moreover, we also observe a redshifted Lyα peak as compared to the other predictions from the literature. This aspect of our predictions is interesting especially in light of the established correlations between emission line shifts, and could be a consequence of its nearest-neighbour quasars in our standardized PCA space or even potentially hint at a new correlation between emission line profiles. Since obtaining a physical basis for machine learning algorithms is challenging, more investigation needs to be done to better understand this aspect.

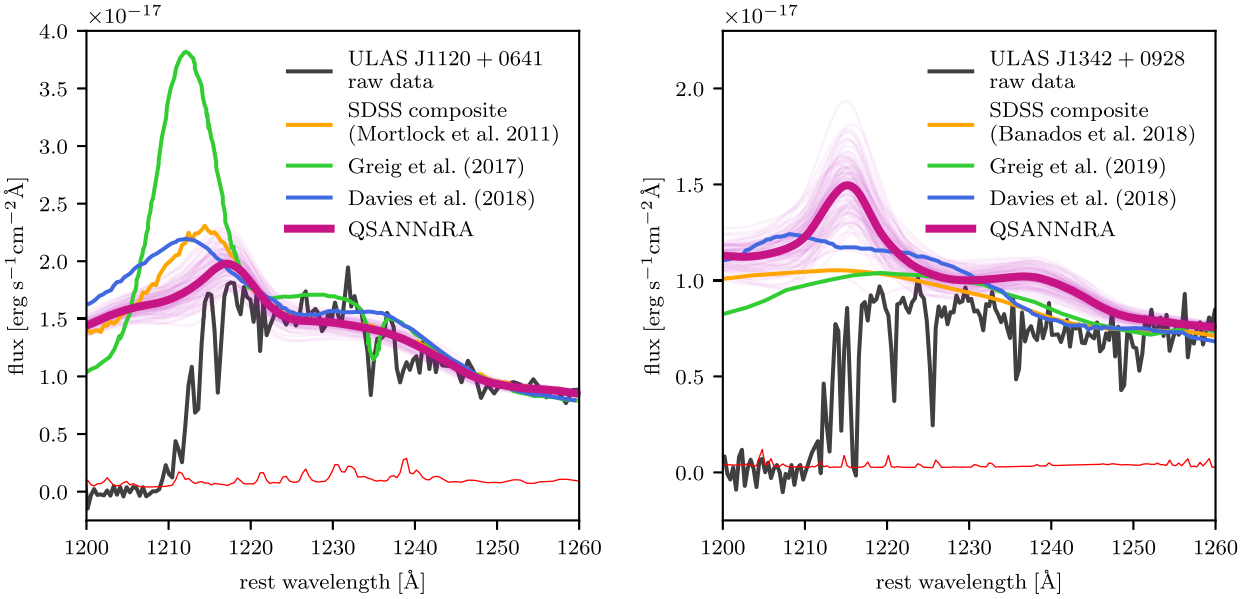

Figure 10. (Left) A comparison of QSANNdRA's prediction (magenta) for the blue-side spectrum of ULAS J1120+0641 with existing predictions from the literature, in particular by Mortlock et al. (2011) (orange), Greig et al. (2017a) (green) and Davies et al. (2018a) (blue). We display raw observational data in grey and their corresponding flux uncertainties in red. (Right) A comparison of QSANNdRA's prediction (magenta) for the blue-side spectrum of ULAS J1342+0928 with existing predictions from the literature, in particular by Bañados et al. (2018) (orange), Greig et al. (2019) (green), and Davies et al. (2018a) (blue). We display raw observational data in grey and their corresponding flux uncertainties in red.

The right-hand panel of Fig. 10 displays our prediction for the blue-side spectrum of ULAS J1342+0928 (magenta) in comparison to the predictions based on the SDSS composite (orange) (Bañados et al. 2018), the covariance matrix approach (green) (Greig et al. 2019), and the PCA method (blue) (Davies et al. 2018a). In contrast to the $z = 7.1$ QSO where our model predicted the weakest emission lines, for this QSO, QSANNdRA predicts both stronger Lyα and stronger NV emission compared to either the SDSS composite or the PCA method from Davies et al. (2018a). Hence, our model is in no way biased to predicting either stronger or weaker emission compared to other models in the literature. Some of the 100 individual NNs predict Lyα emission as weak as that reconstructed using the PCA method, however, all NNs in our model do agree on a significantly stronger NV emission and hence a different spectral shape than those predicted by the SDSS composite or the PCA method. Most interesting is how the predictions for the neutral fraction compare given the systematic differences between our model and those from the literature.

3.2 Constraining the neutral fraction during the Epoch of Reionization

In order to determine the neutral fraction based on our reconstructed spectra, we model the damping wing redward of the Gunn–Peterson trough (Gunn & Peterson 1965) according to the analytical model presented in Miralda-Escudé (1998) combined with the Gunn-Peterson optical depth as defined by Fan et al. (2006). We note that this approach is less sophisticated compared to the methods employed by Davies et al. (2018b) and Greig et al. (2017b, 2019), as we do not use simulation-based models of the local high-density environments and gas inflows or outflows. We use the following cosmological parameters: $h = 0.6766$, $\Omega_m = 0.3111$ and $\Omega_b h^2 = 0.02242$ (Planck Collaboration VI 2018).

The IGM is modelled as homogeneous and neutral for $z_N < z < z_S$, where z_N is the redshift at which reionization is assumed to be complete, and z_S is the redshift corresponding to the end of the near zone of the QSO. For $z < z_N$ the IGM is assumed to be completely ionized. We set $z_N = 6$ as compared to a $z_N = 7$ used by Bañados et al. (2018), however, further analysis showed that the model is largely insensitive to the exact value of this redshift (see Appendix D). Furthermore, the model assumes a fully ionized proximity zone.

We use a common definition for the QSO proximity zone described in the literature (Fan et al. 2006; Carilli et al. 2010; Keating et al. 2015; Venemans et al. 2015; Davies et al. 2018a). We normalized the two fitted high-redshift spectra with respect to the observed flux at Lyα, and defined the end of the proximity zone to correspond to the wavelength at which the fitted flux falls below 10 per cent of the peak value. As an aside, we tested how varying this threshold value impacts the damping wing fit and found that our resultant constraints are insensitive to the exact percentage chosen for the proximity zone definition provided it is <15 per cent. With the blue-side predictions at hand for each $z > 7$ QSO, we then performed a least-squares optimization to constrain the neutral fraction $\bar{x}_{HI}$ in the damping wing model (Miralda-Escudé 1998) by fitting the damped reconstructed continuum to the smoothed continuum estimate in the wavelength range 1210–1250 Å. We used our smoothing algorithm (Appendix B) to avoid fitting to the absorption features in this wavelength range for both QSOs; however, we note that this works better for the $z = 7.5$ quasar than for the lower resolution $z = 7.1$ QSO. The optimization procedure was performed on each prediction within the committee individually to obtain a distribution of possible $\bar{x}_{HI}$ for each QSO. Finally, the resulting values of $\bar{x}_{HI}$ were averaged according to the weights of the individual NNs within the committee to obtain a constraint on the neutral fraction at the redshift of each of the two $z > 7$ QSOs under consideration. These are shown as the blue lines in Fig. 9.

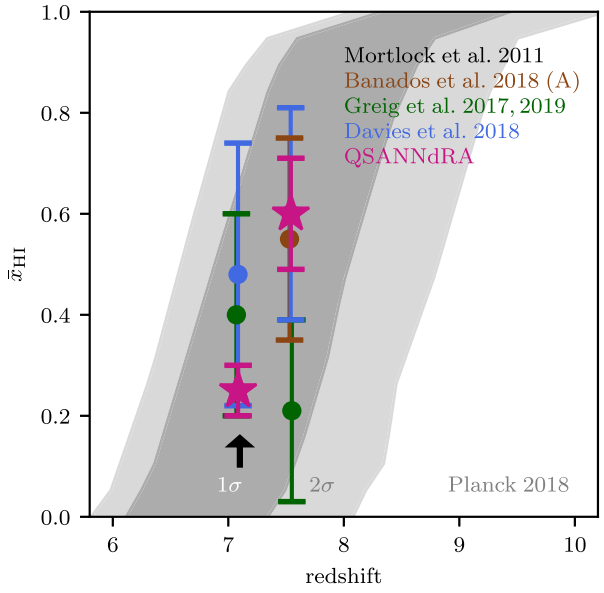

Figure 11. Plot showing the neutral fraction constraints published to date. The dark grey and light grey regions show the 1σ and 2σ constraints, respectively, published by the Planck Collaboration VI (2018) based on CMB observations. We further show constraints based on the damping wing of ULAS J1120+0641 and ULAS J1342+0928 as follows: Mortlock et al. (2011) in black, Greig et al. (2017b, 2019) in green, Bañados et al. (2018) in brown, Davies et al. (2018b) in blue, and finally our constraints in magenta. We emphasize that due to differences in damping wing models, great care should be taken when directly comparing our constraints to those of Davies et al. (2018b) and Greig et al. (2017b, 2019).

In both high-redshift QSOs, our reconstructed spectrum combined with the damping wing model provides a good representation of the observed data.

We specify the 68 per cent confidence bounds on our neutral fraction constraints as the standard deviation of the predictions from the 100 NNs in our ensemble. We find: $\bar{x}_{\rm HI} = 0.25^{+0.05}_{-0.05}$ for ULAS J1120+0641 at $z = 7.0851$, and $\bar{x}_{\rm HI} = 0.60^{+0.11}_{-0.11}$ for ULAS J1342+0928 at $z = 7.5413$.

In Fig. 11, we compare our neutral fraction predictions with the constraints from the Planck Collaboration VI (2018) based on CMB observations as well as the estimates from the other models in the literature that modelled the damping-wing of the two high-redshift QSOs. Our predictions for $\bar{x}_{\rm HI}$ and the 1σ uncertainties at $z = 7.0851$ and $z = 7.5413$ are well within the 1σ contours of the estimated prediction from Planck suggesting good agreement between our method and CMB data. Compared to other models in the literature that modelled the damping-wing, our estimates for $\bar{x}_{\rm HI}$ are comparable at $z = 7.5413$, except for Greig et al. (2019), and tend to fall lower at $z = 7.0851$. This is because for the QSO at $z = 7.0851$, we predict a weaker Lyα emission line compared to these other models which means that less neutral hydrogen is needed in the IGM to account for the observed damping wing.

We emphasize that the constraints on the neutral fraction from the literature that are based on the damping-wing use both a different model for reconstructing QSO spectra as well as a different model for how the spectra are processed by the IGM. We use a simple model of an ionized proximity zone, a completely homogeneous IGM, and a reionization redshift. This is nearly identical to the Model A presented in Bañados et al. (2018). Hence, there is a systematic difference in obtaining neutral fraction estimates even after the QSO spectra are reconstructed (see the difference between the three models presented in Bañados et al. 2018). We therefore emphasize that our neutral fraction error bars represent the errors in modelling the intrinsic spectrum and don't reflect any systematic uncertainties due to the differences between homogeneous and inhomogeneous reionization. Nevertheless, with these differences in mind, our model agrees with the others to within 1σ. In all models, the Universe is neither 100 per cent neutral at $z = 7.5413$ nor is it 100 per cent reionized by $z = 7.0851$. Because of the additional uncertainties in modelling the damping-wing using a more sophisticated model, a significantly larger number of $z > 7$ QSOs along multiple lines of sight will be needed to have a statistical estimate for the high-redshift neutral fraction. Nevertheless, our modelling favours a rapid end to reionization.

4 SUMMARY AND CONCLUSIONS

We have implemented an ensemble of 100 weighted 4-layer fully connected feed-forward NNs termed QSANNdRA for the purpose of reconstructing the intrinsic high-redshift QSO spectra in the damped region around Lyα. We subsequently use these reconstructions to constrain the neutral gas fraction of the IGM at $z > 7$. We trained each individual network in the committee to extract the correlations between the red-side (1290 Å $<\lambda_{\rm rest} <$ 2900 Å) and the blue-side (1192 Å $<\lambda_{\rm rest} <$ 1290 Å) spectral features in a sample of 13 703 quasar spectra at redshifts of $2.09 < z < 2.51$ from the SDSS data base (York et al. 2000; Eisenstein et al. 2011; Dawson et al. 2013; Blanton et al. 2017; Abolfathi et al. 2018). We applied our trained model to two of the highest redshift QSOs known to date, in particular to ULAS J1120+0641 at $z = 7.0851$ (Mortlock et al. 2011; Venemans et al. 2017a), and ULAS J1342+0928 at $z = 7.5413$ (Bañados et al. 2018) to reconstruct their continua around Lyα. Comparison of our model to the state-of-the-art model reported by Davies et al. (2018a) revealed a 14.2 per cent improvement in the mean relative prediction error across previously unseen low-redshift SDSS QSOs. By extending the PCA model to achieve a fairer comparison, we achieved a 6.1 per cent improvement in the mean relative prediction error with the improvement being even more significant for QSOs similar to ULAS J1120+0641 (22.1 per cent improvement) and to ULAS J1342+0928 (16.8 per cent improvement). Finally, we used our predicted continua and a homogeneous reionization model to constrain the volume-averaged neutral fraction at the redshifts $z = 7.0851$ and $z = 7.5413$ to be $\bar{x}_{\rm HI} = 0.25^{+0.05}_{-0.05}$ and $\bar{x}_{\rm HI} = 0.60^{+0.11}_{-0.11}$ (with 68 per cent bounds), respectively.

We emphasize that our constraints use a homogeneous model for the damping wing analysis (Miralda-Escudé 1998), and so our recovered uncertainties on the neutral gas fraction are likely to be underestimates due to a lack of stochasticity coming from inhomogeneous reionization. A much larger sample of observed high-redshift QSOs will be needed to truly understand this effect. None the less, these constraints are consistent with the literature both for ULAS J1220+0641 (Mortlock et al. 2011; Greig et al. 2017b; Davies et al. 2018a) and ULAS J1342+0928 (Bañados et al. 2018; Davies et al. 2018a), as well as the estimates from the CMB (Planck Collaboration VI 2018). However, our predictions lie on the lower end of the existing bounds on the neutral fraction at $z = 7.0851$ compared to other work that modelled the damping-wing. This is because for ULAS J1220+0641, our model predicts weaker intrinsic Lyα emission compared to other models. In addition, the fact that our model predicts a strong N V emission line for

ULAS J1342+0928, which overpredicts the observed flux, affects the interpretation of our neutral fraction constraint.

This is a particularly interesting result, especially in light of the robustness of our prediction model. As Davies et al. (2018b) and others (Mortlock et al. 2011; Bañados et al. 2018) pointed out, both of these QSOs exhibit outlying spectral features as compared to the low-redshift QSO spectra, the most notable feature being the extremely blueshifted C IV line. Our model is able to capture these outlying features well and thus take their full extent into account when making a prediction. Furthermore, based on our trained autoencoder, the red-side spectral features of the $z = 7.1$ QSO are not extreme outliers compared to our training data, and our model is actually expected to perform better on QSOs similar to the ones at $z > 7$ compared to the average QSO in SDSS.

The accuracy of our predictions is particularly interesting for another reason. Even though the normalization and standardization performed in Section 2.2 was done in order to make the various quasar spectra comparable, it also removed the sensitivity to the Baldwin effect (Baldwin 1977). Furthermore, quasar luminosity also correlates with emission line shifts (Shang et al. 2003; Richards et al. 2011) possibly because of physical reasons such as orientation (Meyer, Bosman & Ellis 2019). Despite also removing these correlations from our input data set, our model was still able to achieve low prediction errors, which suggests that these correlations might be implicitly contained in other spectral features.

There are two main strengths of our method in the context of high-redshift QSO continua reconstruction. First, we ensure generalizability of QSANNDRA's performance by constructing the model only on 80 per cent of QSOs (train subset) from our SDSS training set, while we assess its performance on the remaining, previously unseen, 20 per cent of QSOs (test subset). The fact that the difference between the performance of our model on the train and test subsets is small suggests that our model can be applied to other, previously unseen QSO without losing accuracy. Secondly, using artificial NNs is particularly well suited for extracting empirical correlations from large data sets, since it also allows for capturing non-linear relationships among the various spectral features.

Some concerns might arise about the underlying idea of applying the low-redshift spectral correlations to the high-redshift quasars with unusual spectra, however, these remain to be the best resource available to date for the study of intrinsic spectra of these high-redshift objects.

It is also debatable whether modelling the intervening IGM as homogeneously neutral between the end of the near zones of the QSOs and the reionization redshift is reasonable. Reionization is expected to be patchy (e.g. Iliev et al. 2006; Pentericci et al. 2014; Becker, Bolton & Lidz 2015; Bosman et al. 2018; Kulkarni et al. 2019) and therefore we should model a more sophisticated reionization topology than a homogeneous IGM to measure the amount of damping. The highly homogeneous model used in this work (Miralda-Escudé 1998) is clearly in contradiction with this theory. However, due to a lack of a statistical sample of QSOs at redshifts relevant to the Epoch of Reionization, it is difficult to establish the details such a model would require to be accurate. Even though the current use of a simplistic model might be justified this way, it should be emphasized that inhomogeneous models are indeed being employed in other work (Bolton et al. 2011; Greig et al. 2017b; Davies et al. 2018b; Greig et al. 2019). As more high-redshift QSOs are discovered, our very accurate and generalizable QSANNDRA can be used in the context of a more sophisticated damping-wing model to obtain even better constraints on the high-redshift neutral fraction.

ACKNOWLEDGEMENTS

We would like to thank Dr. Daniel Mortlock for providing the observational data for ULAS J1220+0641, and Dr. Eduardo Bañados for the observational data for ULAS J1342+0928 and valuable feedback. The research of JD and AS is partly funded by Adrian Beecroft and STFC.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

REFERENCES

Abolfathi B. et al., 2018, ApJS, 235, 42
Baldwin J. A., 1977, ApJ, 214, 679
Bañados E. et al., 2018, Nature, 553, 473
Becker G. D., Bolton J. S., Lidz A., 2015, Publ. Astron. Soc. Aust., 32, e045
Blanton M. R. et al., 2017, AJ, 154, 28
Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, MNRAS, 416, L70
Boroson T. A., Green R. F., 1992, ApJS, 80, 109
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055
Breiman L., 2001, Mach. Learn., 45, 5
Carilli C. L. et al., 2010, ApJ, 714, 834
Chollet F. et al., 2015, Keras. Available at: https://keras.io
Davies F. B. et al., 2018a, ApJ, 864, 142
Davies F. B. et al., 2018b, ApJ, 864, 143
Dawson K. S. et al., 2013, AJ, 145, 10
Dawson K. S. et al., 2016, AJ, 151, 44
Dietterich T. G., 2000, International Workshop on Multiple Classifier Systems. Springer, Berlin, Heidelberg, p. 1
Eilers A.-C., Davies F. B., Hennawi J. F., Prochaska J. X., Lukić Z., Mazzucchelli C., 2017, ApJ, 840, 24
Eilers A.-C., Davies F. B., Hennawi J. F., 2018, ApJ, 864, 53
Eilers A.-C., Hennawi J. F., Davies F. B., Oñorbe J., 2019, ApJ, 881, 23

Eisenstein D. J. et al., 2011, AJ, 142, 72
Fan X. et al., 2006, AJ, 132, 117
Fischler M. A., Bolles R. C., 1981, Commun. ACM, 24, 381
Francis P. J., Hewett P. C., Foltz C. B., Chaffee F. H., 1992, ApJ, 398, 476
Géron A., 2017, Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc
Greig B., Mesinger A., McGreer I. D., Gallerani S., Haiman Z., 2017a, MNRAS, 466, 1814
Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017b, MNRAS, 466, 4239
Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 5094
Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633
Hewett P. C., Wild V., 2010, MNRAS, 405, 2302
Iliev I. T., Mellema G., Pen U. L., Merz H., Shapiro P. R., Alvarez M. A., 2006, MNRAS, 369, 1625
Jones E., Oliphant T., Peterson P., 2001, SciPy: Open Source Scientific Tools for Python. Available at: http://www.scipy.org/
Keating L. C., Haehnelt M. G., Cantalupo S., Puchwein E., 2015, MNRAS, 454, 681
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Meyer R. A., Bosman S. E. I., Ellis R. S., 2019, MNRAS, 487, 3305
Miralda-Escudé J., 1998, ApJ, 501, 15
Mortlock D., 2016, in Mesinger A., ed., Astrophysics and Space Science Library, Vol. 423, Understanding the Epoch of Cosmic Reionization: Challenges and Progress. Springer, Berlin, p. 187
Mortlock D. J. et al., 2011, Nature, 474, 616
Pâris I. et al., 2011, A&A, 530, A50
Pâris I. et al., 2018, A&A, 613, A51
Pedregosa F. et al., 2011, Journal of machine learning research, 12, 2825
Pentericci L. et al., 2014, ApJ, 793, 113
Planck Collaboration VI, 2018, preprint (arXiv:1807.06209)
Richards G. T. et al., 2011, AJ, 141, 167
Shang Z., Wills B. J., Robinson E. L., Wills D., Laor A., Xie B., Yuan J., 2003, ApJ, 586, 52
Shang Z., Wills B. J., Wills D., Brotherton M. S., 2007, AJ, 134, 294
Shen Y. et al., 2016, ApJ, 831, 7
Stoughton C. et al., 2002, AJ, 123, 485
Suzuki N., 2006, ApJS, 163, 110
Suzuki N., Tytler D., Kirkman D., O'Meara J. M., Lubin D., 2005, ApJ, 618, 592
Venemans B. P. et al., 2015, ApJ, 801, L11
Venemans B. P. et al., 2017a, ApJ, 837, 146
Venemans B. P. et al., 2017b, ApJ, 851, L8
Yip C. W. et al., 2004, AJ, 128, 2603
York D. G. et al., 2000, AJ, 120, 1579

APPENDIX A: S/N CUT ANALYSIS

This appendix reports the impact that different S/N cuts in the cleaning stage can have on the overall performance of our model.

While QSANNDRA has been trained on a low-redshift data set composed of QSOs with S/N $\geq$ 7, we aim to understand how changing this parameter might impact its performance and predictions. In order to investigate this, we created two more training data sets with a S/N cut of 5 and 9, respectively, and then retrained our model on these new data sets while keeping everything else the same. Fig. A1 shows the test set performance of the two new models as well as the baseline performance reported in the main text of this paper. The performance arising from a S/N cut of 5 is shown in orange, the one arising from a S/N cut of 9 is shown in blue, and the performance of the fiducial model, QSANNDRA, is shown in magenta. For each case, we show $\bar{\epsilon}$ as a solid curve and σ_ϵ as a dashed curve.

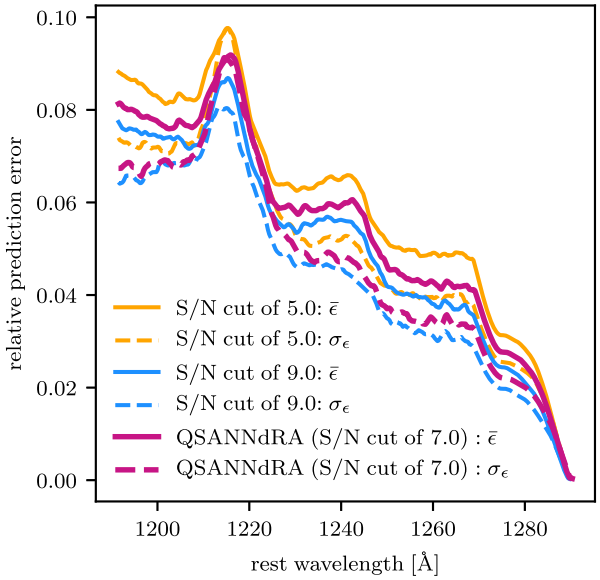

Figure A1. Dependence of QSANNDRA's performance on the S/N cut threshold in the data pre-processing stage shown in terms of $\bar{\epsilon}$ (solid lines) and σ_ϵ (dashed lines). The thick magenta curves display QSANNDRA's performance reported in the main text, while the orange and the blue curves display the model's performance when trained on a data set of low-redshift quasars with S/N $\geq$ 5 and S/N $\geq$ 9, respectively.

While it seems that the performance improves for higher S/N cuts, the conclusion from this analysis has to be drawn carefully. For instance, while it is easier for the networks to learn the underlying spectral correlations in a less noisy data set, using a higher S/N cut comes at the cost of a smaller training data set. In particular, while the original S/N $\geq$ 7 data set has 13 703 quasars, the S/N $\geq$ 5 data set has 22 753 quasars and the S/N $\geq$ 9 has only 9229 quasars. These two aspects ultimately need to be balanced to achieve both a considerably small prediction uncertainty as well as good generalizability.

We further applied the two retrained models to the two high-z quasars, namely ULAS J1120+0641 at $z = 7.0851$ (Mortlock et al. 2011), and ULAS J1342+0928 at $z = 7.5413$ (Bañados et al. 2018), and used their predictions to constrain the neutral fraction as outlined in the main text. Fig. A2 displays the resultant predictions for ULAS J1120+0641 (top panel) and ULAS J1342+0928 (bottom panel) based on the S/N $\geq$ 5 and S/N $\geq$ 9 data sets in orange and blue, respectively, and also shows our main prediction in magenta for comparison. We observe a slight increase in the predicted flux for the $z = 7.1$ QSO and a slight decrease in the predicted flux for the $z = 7.5$ QSO, which then translates into slightly higher neutral fraction constraints for ULAS J1120+0641 and slightly lower neutral fraction constraints for ULAS J1342+0928 (Fig. A3).

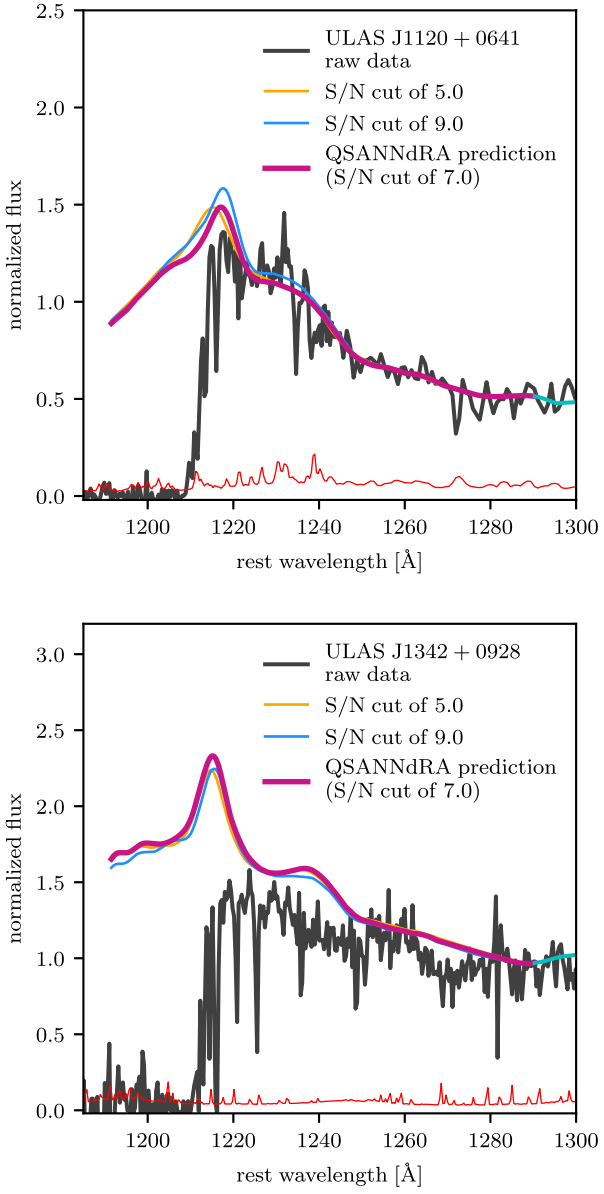

Figure A2. A comparison of the predicted fluxes based on S/N $\geq$ 5 (orange) and S/N $\geq$ 9 (blue) data sets of low-z quasars to the main predictions of QSANNdRA based on a S/N $\geq$ 7 (magenta) data set of low-z quasars for ULAS J1120+0641 (top) and ULAS J1342+0928 (bottom).

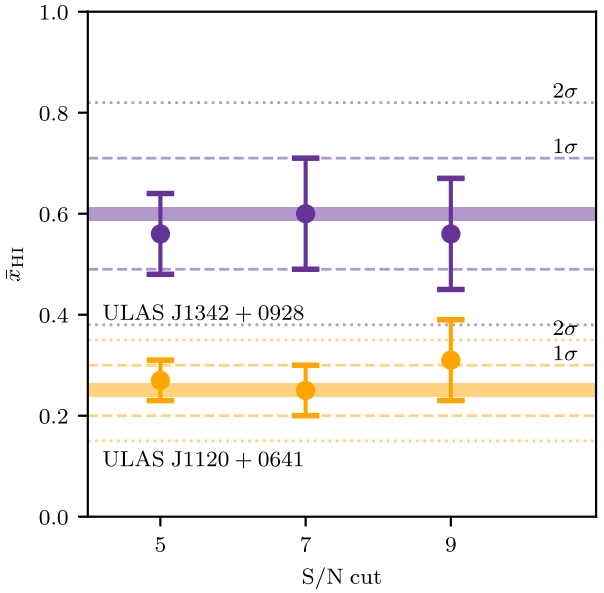

Figure A3. A comparison of the resultant neutral fraction constraints for ULAS J1120+0641 (orange) and ULAS J1342+0928 (purple) arising from a S/N cut of 5, 7, and 9 on the low-redshift SDSS quasars used for the construction of QSANNdRA. The horizontal solid lines depict the baseline neutral fraction constraints $\bar{x}_{\rm HI} = 0.25$ for ULAS J1120+0641 and $\bar{x}_{\rm HI} = 0.60$ for ULAS J1342+0928, while the dashed and dotted lines represent the posterior 1σ and 2σ uncertainty bounds, respectively.

APPENDIX B: SMOOTHING ALGORITHM

This appendix details the smoothing algorithm used for spectral fitting of all quasars in this work. A visual demonstration of the procedure is shown in Fig. B1.

The challenges this algorithm has to overcome are 2-fold. First, we need to compute a fit of the observed flux which discards all absorption features, since these are not part of the *intrinsic* spectra we aim to extract correlations from. Second, we need to capture the full strength of emission line peaks in the spectra without damping them, since these are used to establish the spectral correlations. Based on these two challenges, we developed the following smoothing algorithm which successfully discards all reasonably narrow absorption features while maintaining the full strength of even the strongest and sharpest emission lines, especially the Lyα peak.

We first compute a running median with a bin size of 50 data points in order to capture the main continuum and the overall spectral shape of the spectrum. This curve then acts like a median border [Fig. B1 (a), yellow], above which we then perform a peak-finding algorithm using the SCIPY PYTHON library (Jones et al. 2001) to find local maxima in the spectrum [Fig. B1 (a), red points].

We then interpolate the peaks to construct an upper envelope of the spectrum [Fig. B1 (b), red]. This envelope is then subtracted from the raw data points, thus linearizing our data into residuals [Fig. B1 (c), black]. We then apply the RANSAC regressor algorithm (Fischler & Bolles 1981) from the SCIKIT-LEARN PYTHON package (Pedregosa et al. 2012) on the residuals, which fits a linear function to the data and masks all data points as either inliers or outliers. By further considering only inlying data points [Fig. B1 (c), green], we thus reject most absorption features in the spectrum.

Finally, the data points which are flagged as inliers by RANSAC are interpolated and smoothed by computing a running median

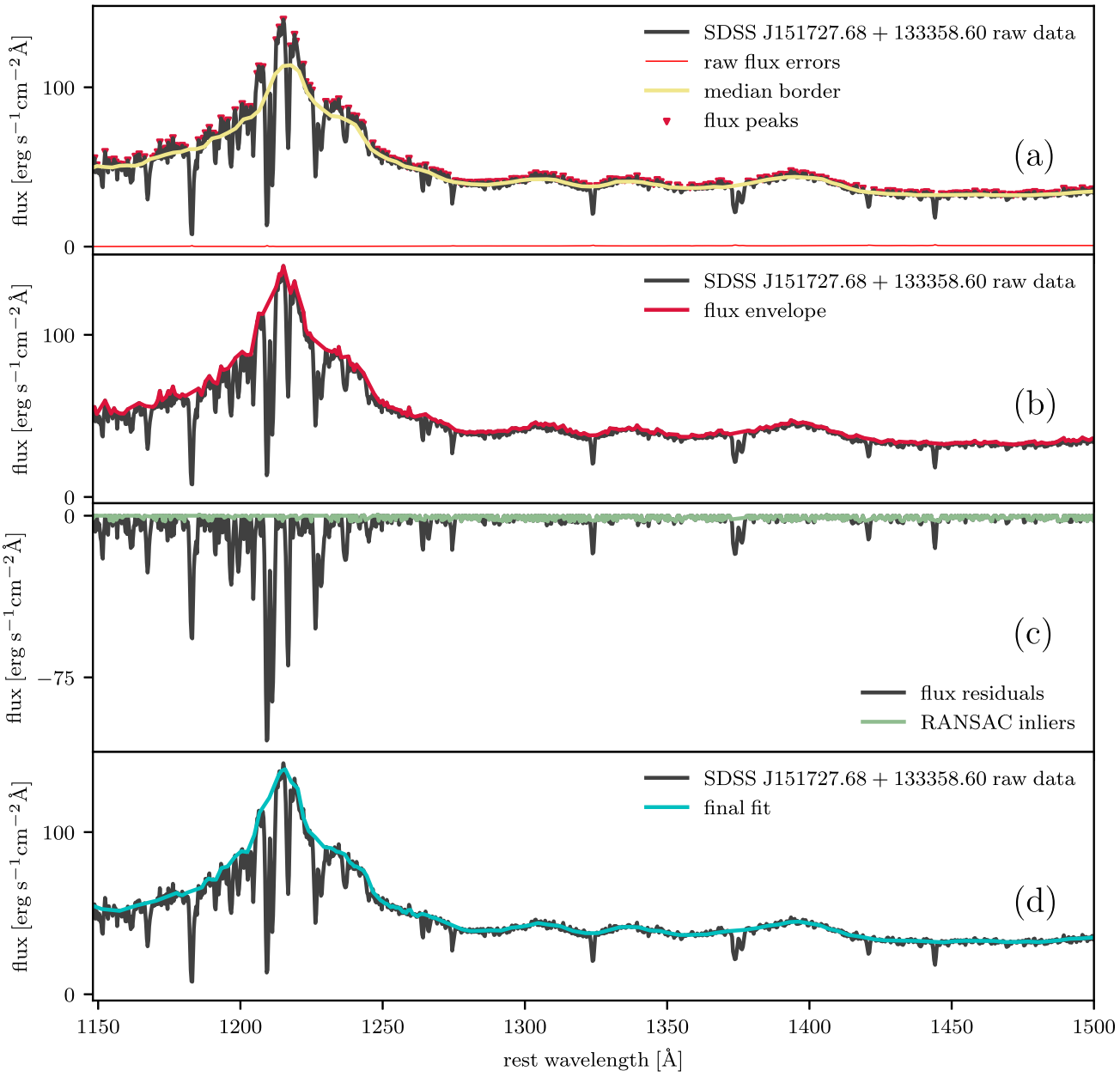

Figure B1. An illustration of our smoothing algorithm. We first compute a running median with a bin size of 50 data points to capture the main continuum and emission features in the spectrum (a, yellow). We then perform a peak-finding procedure using the SCIPY PYTHON library (Jones et al. 2001) above the aforementioned running median border (a, red points) and interpolate the peaks to construct and upper envelope of the spectrum (b, red). This envelope is then subtracted from the spectrum, thus linearizing our data (c, black). We then applied the RANSAC regressor algorithm (Fischler & Bolles 1981) from the SCIKIT-LEARN PYTHON package (Pedregosa et al. 2012) on the residuals, thus rejecting most absorption features in the spectrum. The data points which were flagged as inliers by RANSAC (c, green) were interpolated and smoothed by computing a running median with a bin size of 20, thus creating the final flux fit of each spectrum (d, cyan).

with a bin size of 20, thus creating the final flux fit of each spectrum [Fig. B1 (d), cyan]. This algorithm will be made public in near future.

APPENDIX C: ANALYSIS OF REDSHIFT CALIBRATION SYSTEMATICS

This section discusses and analyses an important potential source of systematic errors, namely the uncertainty in the QSO redshifts, both the low-redshift ones (i.e. training set SDSS quasars) and the high-redshift ones (i.e. ULAS J1120+0641 and ULAS J1342+0928). Moreover, while the SDSS redshifts have been calculated based on the broad UV emission lines, the redshifts of the two $z > 7$ QSOs come from the sub-millimetre emission from their host galaxies (Venemans et al. 2017a, b; Bañados et al. 2018). We analyse how resampling the redshifts of the SDSS QSOs in the training set influences the performance and predictions of QSANNDRA, and we also investigate how changing the redshifts of the high-redshift QSOs influences the predicted continua and hence neutral fractions. As a final test, we recalibrate both the low-redshift and the high-redshift spectra based on the redshift coming from the Mg II line,

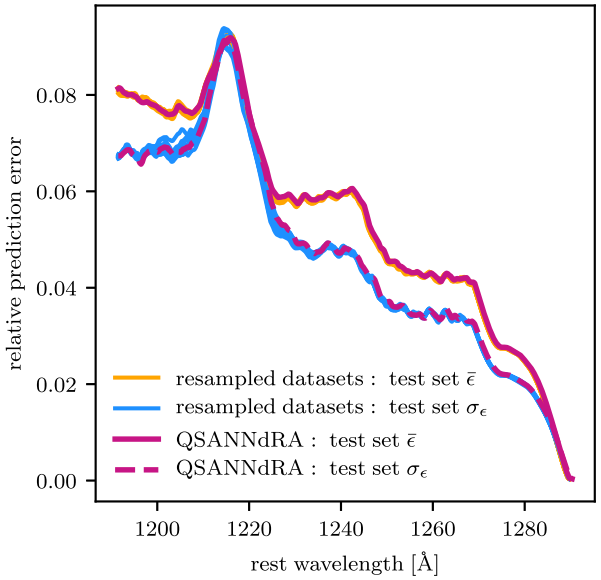

Figure C1. Wavelength dependence of the test set performance of our model based on the 10 resampled data sets displayed as $\bar{\epsilon}$ (orange) and σ_ϵ (blue). Both of these metrics are almost identical to the baseline performance from the main text (magenta), hence showing that redshift calibration of the training set quasars does not influence the performance of our model.

since it has been shown to be the least affected by systematic shifts (Hewett & Wild 2010; Shen et al. 2016).

The first part of our analysis involves investigating how changing the redshifts of the SDSS QSOs and retraining our model changes the performance and predictions for the two high-redshift QSOs. We do this by resampling our training set QSOs 10 times to create 10 different training sets. For each new data set, noting that each SDSS QSO has a redshift uncertainty Z_ERR assigned to it which defines the variance of its redshift distribution, we randomly sample the redshift of each QSO along this distribution and calibrate the observed wavelengths to the correct rest frame accordingly. In doing this, we model the redshift distribution of each QSO as Gaussian with a mean of Z and a sigma of Z_ERR. With these 10 resampled training sets, we proceed in the exact same way as reported in the main text, i.e. we perform a 10-fold cross-validation to clean up the training sets, we train QSANNdRA on these new data sets, and we predict high-redshift quasar continua along with neutral fraction constraints.

After training, we evaluate the performance of our model on the test set of low-redshift QSOs by means of the mean absolute prediction error $\bar{\epsilon}$ and its standard deviation σ_ϵ. Fig. C1 displays QSANNdRA's performance on each of the 10 resampled data sets as $\bar{\epsilon}$ in orange and σ_ϵ in blue, along with its performance from the main text in magenta. As can be clearly seen, the differences in the performances are marginal and hence we conclude that redshift calibration does not significantly influence the performance our model is able to achieve, if the errors are distributed randomly. Note that this may change if the redshifts are systematically biased in any way for certain types of QSOs. Given the complexity of our NNs, we do expect that our training should be able to account for some of this systematic bias if it is due to the redshift estimation from specific emission lines.

With the 10 new trained models, we reconstruct the blue-side continua for both ULAS J1120+0641 and ULAS J1342+0928. Fig. C2 shows all 10 resultant predictions for ULAS J1120+0641 (top) and ULAS J1342+0928 (bottom), respectively, as a set of

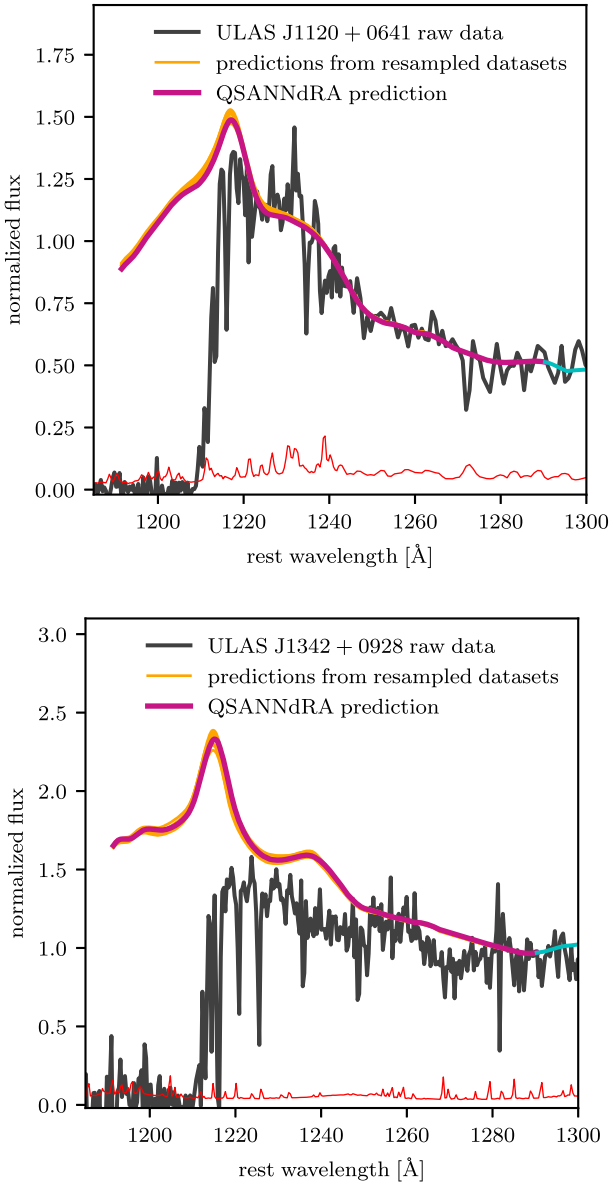

Figure C2. A comparison of the 10 new predictions (orange curves) based on the resampled data sets and QSANNdRA's baseline prediction for ULAS J1120+0641 (top) and ULAS J1342+0928 (bottom). All predictions almost completely overlap, the most significant exception being the Lyα peak where the variance is the greatest, yet still considerably small. Even though this influences the resultant neutral fraction constraint, our model seems to be robust against changes in the redshift calibration of the SDSS training set QSOs.

orange lines and the predictions given in the main text in magenta for comparison. In each case, all resultant predictions from the resampled data sets almost completely overlap with our baseline prediction, which hints at a very marginal influence of redshift calibration on the predicted continua themselves. However, it should be noted that the largest spread in predictions occurs at the Lyα peak in both cases, which can in turn influence the predicted neutral fraction constraints.

The resultant neutral fractions computed by the 10 retrained models are displayed in Fig. C3 along with their corresponding 68 per cent bounds for both ULAS J1120+0641 (orange) and ULAS J1342+0928 (purple). We also show the original predictions

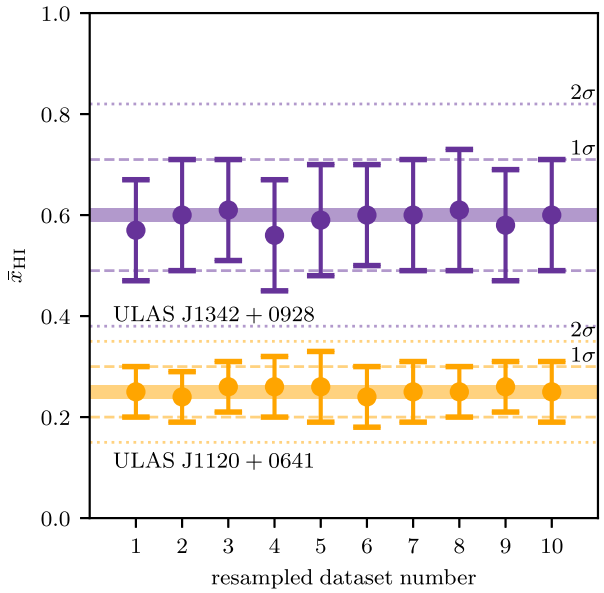

Figure C3. A comparison of the predicted neutral fractions based on the 10 resampled data sets for ULAS J1120+0641 (orange) and ULAS J1342+0928 (purple), to the baseline predictions from the main text (thick horizontal lines). All neutral fraction constraints based on the resampled data sets are consistent with each other as well as with the baseline constraints $\bar{x}_{\rm HI} = 0.25^{+0.05}_{-0.05}$ for ULAS J1120+0641 and $\bar{x}_{\rm HI} = 0.60^{+0.11}_{-0.11}$ for ULAS J1342+0928.

from the main text as grey horizontal lines as well as their corresponding 1σ (dashed) and 2σ (dotted) bounds. All 10 predictions are consistent with each other and also with the fiducial prediction in each case.

The second part of this analysis investigates how QSANNDRA's predictions change as we vary the redshift of the two high-redshift QSOs. Venemans et al. (2017a) reported a redshift of $7.0851^{+0.0005}_{-0.0005}$ for ULAS J1120+0641, and Venemans et al. (2017b) and Bañados et al. (2018) reported a redshift of $7.5413^{+0.0007}_{-0.0007}$ for ULAS J1342+0928. Hence, we re-run our fiducial model on these two QSOs again, each time changing the redshift based on which calibration from observed to rest-frame wavelengths was carried out. Fig. C4 shows the resultant neutral fractions as blue data points with errorbars corresponding to 68 per cent bounds predicted by QSANNDRA. The reported redshift and 1σ and 2σ bounds are shown as dashed and dotted grey lines, respectively.

As can be observed, the relationship between the redshift of both high-redshift QSOs and the estimated $\bar{x}_{\rm HI}$ seems to be approximately linear and is likely to be the consequence of all the spectral features being translated along the wavelength space, thus changing the values of red-side PCA coefficients that QSANNDRA is basing its predictions on. Within the 2σ uncertainty in redshift quoted for the both QSOs, there is virtually no change in the estimated neutral fraction.

As the final test, we used redshifts based on the Mg II line to recalibrate both the SDSS and $z > 7$ spectra in a unified fashion and then retrained our model to predict the high-redshift continua and neutral fraction constraints. This particular line was chosen due to its minimal systematic shifts (Hewett & Wild 2010; Shen et al. 2016). For the low-redshift QSOs, we used the Mg II redshift from the SDSS pipelines, while for the high-redshift quasars, we estimated the corresponding redshift based on the Mg II peak wavelength of the fitted continuum. However, since the sub-millimetre redshifts ($z = 7.0851$ and $z = 7.5413$) for the high-redshift QSOs are physically much more accurate than an estimate from the Mg II emission line, we perform the damping wing analysis in the rest frame of these quasars defined by z.

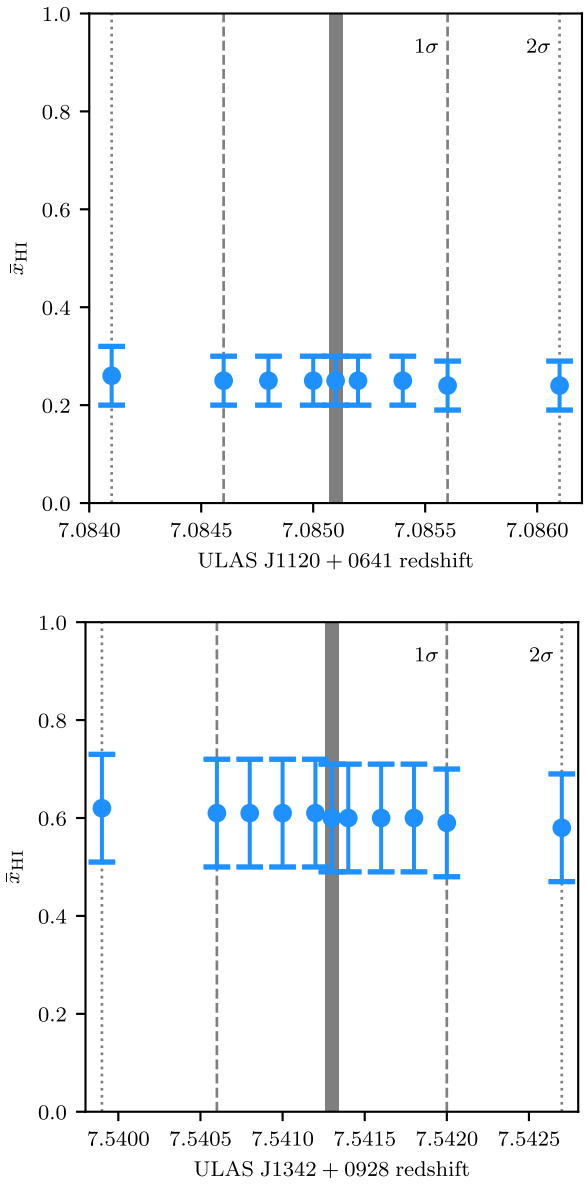

Figure C4. Dependence of the estimated neutral fraction $\bar{x}_{\rm HI}$ predicted by QSANNDRA for ULAS J1120+0641 (top) and ULAS J1342+0928 (bottom) on the redshift of the QSO. The vertical, solid grey line depict the reported redshifts, while the dashed and dotted lines represent the 1σ and 2σ uncertainties, respectively. The error bars on the blue points represent the 68 per cent confidence interval on the estimate of the neutral fraction.

In Fig. C5, we show the resultant predictions and neutral fraction constraints for both $z > 7$ QSOs. We observe that there is a minimal change in the shape of both predicted continua as well as the strength of the Lyα peak. Note that the y-axis in both plots has been normalized with respect to the fitted flux at 1290 Å, which corresponds to a different value than that in the main text. In addition, the new neutral fraction constraints, namely $\bar{x}_{\rm HI} = 0.24^{+0.06}_{-0.06}$ at $z = 7.0851$ and $\bar{x}_{\rm HI} = 0.53^{+0.13}_{-0.13}$ at $z = 7.5413$, are consistent with the constraints from the main text ($\bar{x}_{\rm HI} = 0.25^{+0.05}_{-0.05}$ and $\bar{x}_{\rm HI} = 0.60^{+0.11}_{-0.11}$, respectively). We therefore conclude that the systematics due to

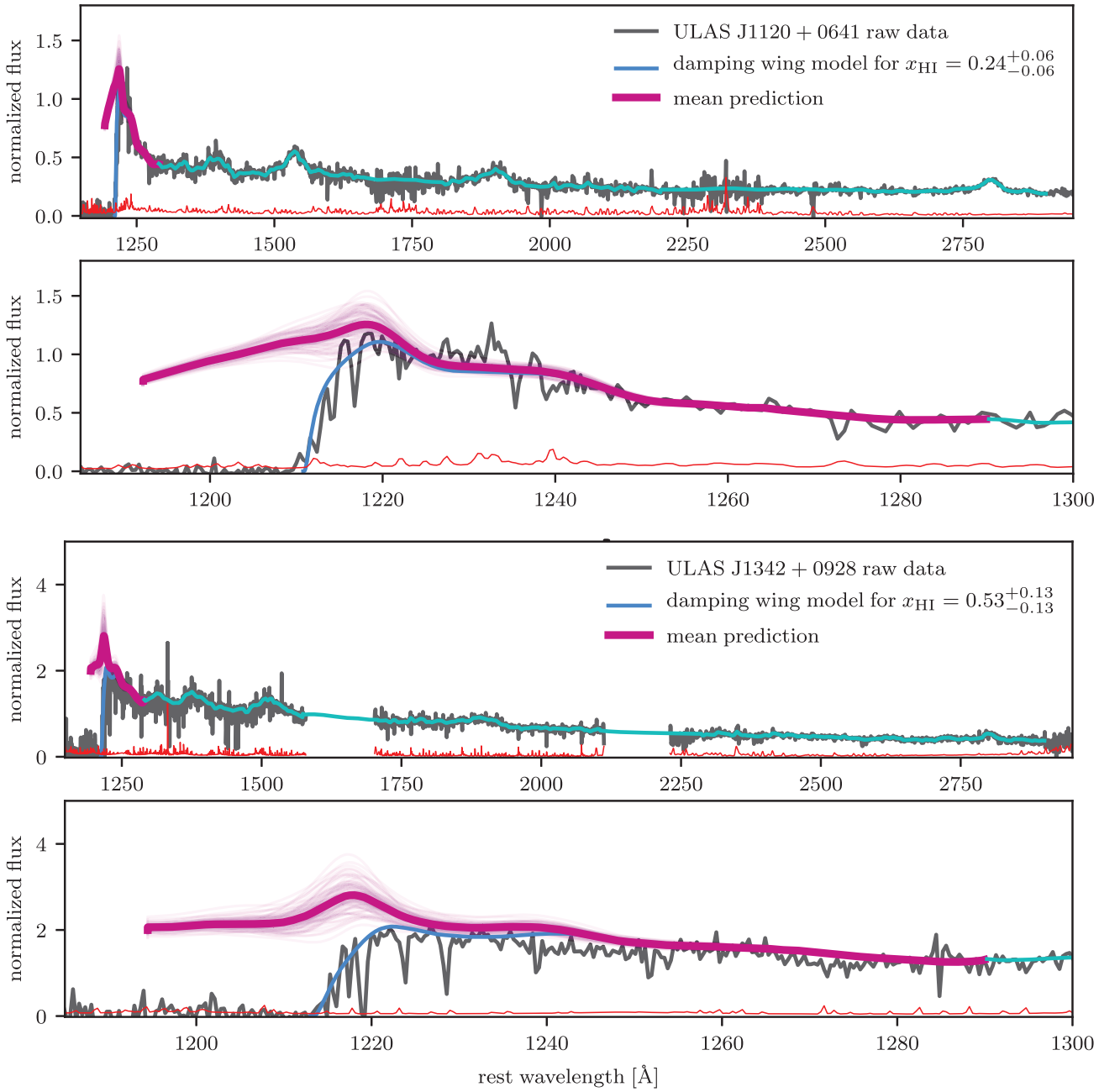

Figure C5. The reconstructed spectrum of ULAS J1120+0641 (top two panels) and ULAS J1342+0928 (bottom two panels) based on a test model recalibrated according to the Mg II emission line redshift. The bottom panel for each quasar shows a close-up view of the Lyα region. The raw data points and their uncertainties are shown in grey and red, respectively. The cyan curve represents our fit of the red-side spectrum. The thin light magenta lines show the individual predictions from the 100 retrained NNs in the committee, while the thick magenta line shows the weighted average of these predictions at each wavelength. The damping wing model is shown in blue and corresponds to $\bar{x}_{\rm HI} = 0.24$ for ULAS J1220+0641 and $\bar{x}_{\rm HI} = 0.53$ for ULAS J1342+0928, which was calculated as the weighted average of optimal neutral fractions corresponding to the individual predictions of the 100 networks within the committee. Note that while the predictions are based on estimates of $z_{z_{\rm MgII}}$ from the fitted spectrum of the QSOs, the damping wing analysis was performed for the much more accurate redshifts, i.e. $z = 7.0851$ and $z = 7.5413$, respectively.

different redshift calibrations do not cause significant errors in our model.

APPENDIX D: DEPENDENCE OF NEUTRAL FRACTION CONSTRAINTS ON $z_{\rm N}$

Here, we report an analysis of how the exact choice of the $z_{\rm N}$ parameter in the damping-wing model (Miralda-Escudé 1998), which defines the redshift by which the IGM is completely reionized, impacts the predicted neutral fraction constraints for the two high-redshift quasars, ULAS J1120+0641 and ULAS J1342+0928.

In this analysis, we use the fiducial QSANNDRA algorithm as described in the main text. We re-run the prediction algorithm multiple times, each time with a different value of $z_{\rm N}$ ranging from $z = 5.5$ up to the redshift of the particular quasar while keeping everything else constant. Fig. D1 shows the resultant neutral

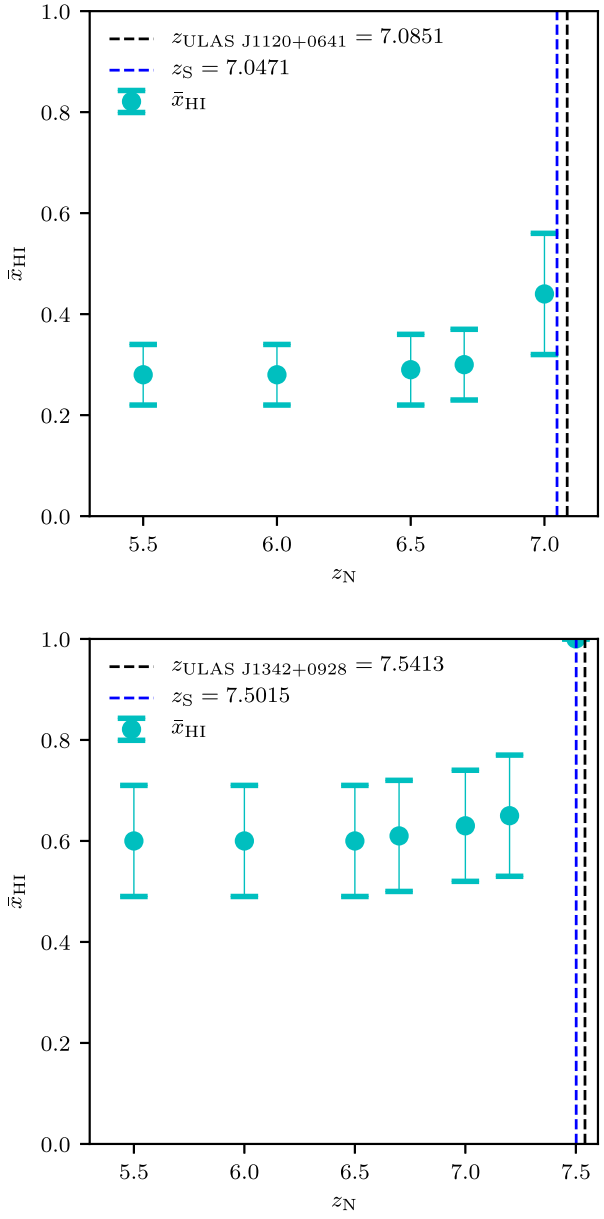

Figure D1. Dependence of the predicted neutral fraction $\bar{x}_{\rm HI}$ for ULAS J1120+0641 (top) and for ULAS J1342+0928 (bottom) on the value of the redshift $z_{\rm N}$, by which we assume the IGM to be fully ionized. We confirm that this dependence is very weak unless $z_{\rm N}$ nears the redshift corresponding to the end of the quasar's proximity zone in the model used in the main text.

fractions and their 68 per cent bounds for ULAS J1120+0641 and ULAS J1342+0928.

In each case, we confirm that the exact value of $z_{\rm N}$ does not have a significant impact on the predictions provided it is not nearing the redshift corresponding to the end of the quasar's near zone $z_{\rm S}$. This makes physical sense since the difference between $z_{\rm N}$ and $z_{\rm S}$ constrains the distance range over which the observed damping by neutral hydrogen in the IGM needs to happen. If this range gets extremely small, the neutral fraction needed to reconstruct the observed damping-wing must increase, which is what we see in the last few data points nearing the QSO redshifts in Fig. D1.

Overall, these results confirm that our choice of $z_{\rm N} = 6$ is not particularly important and that our main results can be generalized well to values of $z_{\rm N} \lesssim 6.5$.

This paper has been typeset from a TEX/LATEX file prepared by the author.