Quasar Factor Analysis—An Unsupervised and Probabilistic Quasar Continuum Prediction Algorithm with Latent Factor Analysis

Zechang Sun (孙泽昌)1, Yuan-Sen Ting (丁源森)2,3, and Zheng Cai (蔡峥)1,4

1 Department of Astronomy, Tsinghua University, Beijing 100084, People’s Republic of China; szc22@mails.tsinghua.edu.cn
2 Research School of Astronomy & Astrophysics, Australian National University, Cotter Road, Weston, ACT 2611, Australia
3 School of Computing, Australian National University, Acton, ACT 2601, Australia
4 Department of Mathematics and Theories, Peng Cheng Laboratory, Nanshan, Shenzhen, People’s Republic of China

Received 2022 November 18; revised 2023 August 11; accepted 2023 August 15; published 2023 October 23

Abstract

Since their first discovery, quasars have been essential probes of the distant Universe. However, due to our limited knowledge of its nature, predicting the intrinsic quasar continua has bottlenecked their usage. Existing methods of quasar continuum recovery often rely on a limited number of high-quality quasar spectra, which might not capture the full diversity of the quasar population. In this study, we propose an unsupervised probabilistic model, quasar factor analysis (QFA), which combines factor analysis with physical priors of the intergalactic medium to overcome these limitations. QFA captures the posterior distribution of quasar continua through generatively modeling quasar spectra. We demonstrate that QFA can achieve the state-of-the-art performance, ∼2% relative error, for continuum prediction in the Lyα forest region compared to previous methods. We further fit 90,678 2 < z < 3.5, signal-to-noise ratio >2 quasar spectra from Sloan Digital Sky Survey Data Release 16 and found that for ∼30% quasar spectra where the continua were ill-determined with previous methods, QFA yields visually more plausible continua. QFA also attains ≲1% error in the 1D Lyα power spectrum measurements at z ∼ 3 and ∼4% in z ∼ 2.4. In addition, QFA determines latent factors representing more physical motivation than principal component analysis. We investigate the evolution of the latent factors and report no significant redshift or luminosity dependency except for the Baldwin effect. The generative nature of QFA also enables outlier detection robustly; we showed that QFA is effective in selecting outlying quasar spectra, including draped Lyα systems and potential Type II quasar spectra.

1. Introduction

Powered by the accretion of matter into supermassive black holes in the galactic nuclei, luminous quasars can shine across vast cosmic distances. As such, not only are they interesting astronomical objects that can reveal the enigmatic physics about active galactic nuclei (AGNs; e.g., Shen & Ho 2014), they also serve as light beacons and shed light on topics that would otherwise not be accessible to us. The Sloan Digital Sky Survey (SDSS; e.g., Lyke et al. 2020) has been the singular powerhouse in the study of quasars. More than 750,000 quasars have been characterized by the SDSS to date with redshift up to ∼7.5, spanning almost the entire cosmic history. Since quasars can reside in both the near and far Universe, the absorption features in quasar spectra are telltale witnesses of the evolution of the Universe.

The Lyα forest imprinted on the quasar spectra (e.g., Bahcall & Goldsmith 1971; Lynds 1971), damped Lyα absorption systems (e.g., Wolfe et al. 2005), and Gunn–Peterson trough (e.g., Gunn & Peterson 1965; Becker et al. 2001) can all provide critical information about the physical state and matter distribution of the intergalactic medium (IGM). For instance, the measurement of baryonic acoustic oscillations (BAOs) from the Lyα absorption forest (e.g., Slosar et al. 2013; Font-Ribera et al. 2014; du Mas des Bourboux et al. 2020) has been one of the mainstream methods to constrain cosmological parameters. The Lyα forest power spectrum can be used to infer the temperature distribution of the IGM at high redshift (e.g., Palanque-Delabrouille et al. 2013; Chabanier et al. 2019) and reveal critical information about the epoch of reionization and cosmic dawn (e.g., Montero-Camacho & Mao 2020). The Gunn–Peterson damping wing signatures in z ≥ 6 quasar spectra is a direct probe to constrain the reionization history (e.g., Davies et al. 2018a; Greig et al. 2019; Šurovčíková et al. 2020). Moreover, cross-correlating the Lyα forest with the Lyα emission can be used to resolve the diffuse Lyα emission on cosmological scales, which unravels the nature of galactic evolution and outflow in the early Universe (e.g., Croft et al. 2018; Renard et al. 2020; Lin et al. 2022).

However, how well we can use quasars to study the intervening gas between us and the quasars critically depends on how well we can infer the intrinsic background quasar continua. Despite its central role, deriving robust quasar continua, especially for noisy, low-resolution, and high-redshift quasar spectra, has proven to be nontrivial even for well-trained astronomers (e.g., Kirkman et al. 2005; Faucher-Giguère et al. 2008). Our inability to accurately infer the continua has always been a significant source of uncertainty for using quasars as cosmological probes, especially for high-order statistics. For example, Lee (2012) demonstrated that ∼2% continuum prediction uncertainties could double the uncertainty in the study of the temperature–density relation (Hui & Gnedin 1997) of the IGM. Furthermore,
the current uncertainties in continuum determination can typically lead to $\sim 2\%$--$6\%$ bias in the measurements of Ly$\alpha$ power spectrum, (e.g., Palanque-Delabrouille et al. 2013; Chabanier et al. 2019). Ongoing and upcoming large-scale spectroscopic sky surveys, e.g., the Dark Energy Spectroscopic Instrument (DESI; Schlegel et al. 2022), typically acquire $\geq 10^6$ quasar spectra or more (e.g., Chaussidon et al. 2021, 2023). The goal is to achieve cosmological measurements to a percent level of uncertainty (e.g., Karačayev et al. 2020). As we further shrink the statistical errors via a larger sample, the high-precision requirement hinges on our ability to infer the quasar continuum accurately.

Various methods have been proposed to infer the intrinsic quasar continua. A majority of the methods focuses on extracting information from wavelengths redder than the Ly$\alpha$ emission (hereafter, “red-side”), which is largely devoid of the IGM absorption, to infer the continuum at wavelengths bluer than the Ly$\alpha$ emission line (hereafter, “blue-side”). Among the most well-adopted methods, the power-law extrapolation methods (e.g., Fan et al. 2006; Yang et al. 2020) assume the quasar continuum to be a power-law function and extrapolate the fitting from the red side to the blue side. The principal component analysis (PCA)-based methods (e.g., Suzuki et al. 2005; Pâris et al. 2011; Lee et al. 2012; Davies et al. 2018b; Durovčíková et al. 2020) assume linear combinations of a few components can well represent quasar continua. The myriad PCA-based continuum fitting methods mainly differ in the determination of the weight of each linear component, including projection (Suzuki et al. 2005; Pâris et al. 2011; Davies et al. 2018b), least-squares fitting (Lee et al. 2012), and neural networks (Durovčíková et al. 2020). Apart from the mainstream PCA-based approaches, other continuum fitting methods include (1) PICCA continuum fitting (du Mas des Bourboux et al. 2020), which fits for a polynomial correction to the mean quasar continuum to approximate various quasar continua in the blue side, and (2) deep-learning-based methods (e.g., Reiman et al. 2020; Liu & Bordoloi 2021), which directly learn a high-dimensional mapping between the red-side continua and the blue-side continua through neural networks, trained on mock data sets or a subsample of high-quality quasar spectra.

However, these existing methods come with several limitations. The supervised learning methods where we learn how to map the red-side continua to the blue-side continua—including the PCA-based and the deep-learning-based methods—rely on a small portion ($\leq 2\%$) of high signal-to-noise ratio ($S/N$) quasar spectra compared to the whole observations (100%). The training sets typically comprise ad hoc continua derived from hand fitting, (e.g., Suzuki et al. 2005; Pâris et al. 2011), or some automatic smoothing algorithms (e.g., Davies et al. 2018b; Reiman et al. 2020; Durovčíková et al. 2020; Bosman et al. 2021). These methods may be unable to generalize for the vast population of quasar spectra, as the training sample is, intrinsically, biased toward brighter quasars. As for the existing unsupervised learning method, where the task is to generalize the red-blue connection through a parametric model—including the power-law model and PICCA—these parametric models might lack the flexibility that hinders their performance in real-life application.

In light of these limitations, we propose a new unsupervised learning algorithm, quasar factor analysis (QFA) to infer the posterior distribution of the intrinsic quasar continua using the entire observed spectrum instead of just harnessing the red-side information. At its core, QFA aims to model the full joint distribution of all quasar spectra by learning the distribution of quasar continua and the Ly$\alpha$ forest. As we will demonstrate, flexible unsupervised generative models with sufficient physical priors can capture the posterior distribution of quasar continua without ad hoc intervention. Since the generative task only maximizes the likelihood of the data, QFA can naturally harness information from all quasar spectra collected, taking into account their heteroskedastic noises.

The generative nature of QFA further leads to a few other advantages compared to existing methods. First, QFA provides a robust posterior distribution of the quasar continua, which is handy for incorporating the continuum uncertainties into other downstream Bayesian inferences, see Figure 1 for details. Second, unlike PCA, latent factor analysis allows for a more flexible basis for the quasar continuum decomposition, leading to a more robust physical interpretation of the basis. Finally, as a fully probabilistic model of quasar spectra, missing pixels or outlying features (such as strong Ly$\alpha$ absorbers) in the quasar spectra can be rigorously dealt with through marginalizing over them.

This paper is organized as follows: In Section 2, we will discuss the core idea and the probabilistic framework of our unsupervised learning algorithm. In Section 3, we present the quasar spectra studied in this paper, including both the SDSS DR16 quasar spectra and the mock spectra with realistic absorption fields. In Section 4, we will compare the performance of our model with existing methods and further investigate how continuum fitting affects the 1D Ly$\alpha$ forest power spectrum measurements. Additionally, we will use QFA for out-of-distribution detection and study the quasar population’s redshift evolution and luminosity dependency. Finally, we discuss the application of QFA for $z > 5$ quasars, its impact on Bayesian cosmology measurements, its limitations, and future directions in Section 5. We conclude in Section 6.

2. Method

The basic premise of QFA is to build a statistical model to bridge the gap between the observables and the unobservables. Specifically, the quasar continua and the Ly$\alpha$ forest are not directly observable. Only when they are combined can we observe the quasar spectra. QFA aims to independently model the distribution of both quasar continua and transmission fields and integrate them to obtain the marginal distribution of observed quasar spectra. By maximizing the marginal likelihood of quasar spectra, QFA will simultaneously learn the distribution of quasar continua and the transmission fields. The optimized model then allows us to infer the posterior distribution of the quasar continua, conditioning on the observed quasar spectra.

In the following, we will introduce QFA, how to train this model with maximal likelihood estimation, and how to infer the posterior distribution of the quasar continua. Figure 2 demonstrates a schematic summary of how QFA works.

2.1. Latent Factor Analysis

QFA models the quasar continua based on latent factor analysis, hence the name. We will briefly summarize the basics of latent factor analysis to better orient the readers. For interested readers, we will refer to Bartholomew et al. (2011) and Beaujean & Loehlin (2017) for details. We denote all random variables in this paper in boldface and other deterministic variables otherwise. The model parameters we optimize through maximum likelihood are italicized.
Latent factor analysis (Bartholomew et al. 2011; Woods & Edwards 2011; Barber 2012; Beaujean & Loehlin 2017) is a powerful statistical model that assumes that a high-dimensional correlated data set (here, the quasar continua) can be expressed as linear combinations of a set of lower-dimensional latent factors. Formally, the observed data $X$ is assumed to be

$$X = \mu + FH + \Psi,$$

where $\mu$ denotes the mean vector with size $N_s$; $h$ denotes the latent factors with size $N_h$ ($N_h \ll N_s$), $F$ denotes the factor loading matrix with size $N_s \times N_h$, which signifies all of the latent factors, and $\Psi$ denotes the unaccounted “error” term with size $N_s$.

Factor models assume the latent factor $h$ and the $\Psi$ follow Gaussian distributions. Since the factor $h$ is only determined up to a rotation, as in any linear combination of $h$ is interchangeable with the factor loading matrix $F$, $h$ is preset to follow a multivariate normal distribution as $h \sim N(0, I)$. The heterogeneous error term, $\Psi$, accounts for the residual of the data set that the latent factors cannot explain. $\Psi$ is assumed to be distributed as $\Psi \sim N(0, \text{diag} (\sigma^2_{\Psi}))$, where $\sigma^2_{\Psi}$ is a free parameter vector to be optimized for.
Generally, factor analysis can be viewed as a generalized version of PCA. The factor loading matrix $F$ plays the same role as the basis matrix in PCA, and the factor $h$ works the same as the PCA coefficients. However, there are two key differences between a factor model and a PCA model. First, PCA assumes an isotropic error term that is distributed as $\mathcal{N}(0, \sigma^2 I)$, while the error term $\Psi$ in latent factor analysis has an anisotropic covariance matrix $\text{diag}(\sigma^2)$. Second, unlike PCA, which requires a set of orthogonal basis, factor analysis does not necessitate the linear components to be orthogonal. The anisotropic stochastic “error” term and nonorthogonal linear components allow factor analysis more flexibility to adapt to real-life data and better interpretability. Furthermore, the statistical foundation of factor analysis, based on maximum likelihood estimation, makes it possible to construct a complex probabilistic spectrum model from this base model, including other physical priors, which we will explain next. The physical prior is critical in ensuring that the factor model focuses only on extracting the continuum, breaking the degeneracy between the transmission fields and the quasar continua.

### 2.2. QFA—A Generative Model of Quasar Spectra

Building upon the basic latent factor model, we leverage our physical priors on the Ly$\alpha$ forest to construct a full generative model of the quasar spectra. For ease of discussion, throughout this paper, we split the quasar spectra into the red side and the blue side relative to the Ly$\alpha$ emission line (rest-frame wavelength $\lambda_{RF} \approx 1215.67$ Å) to distinguish various variables and their roles. We denote the red-side variables with a subscript $r$ and the blue-side variables with a subscript $b$.

We model the quasar continua $C$ through a factor model:

$$C = \mu + Fh + \Psi.$$  \hspace{1cm} (2)

The quasar spectra are further modified for the blue side by the Ly$\alpha$ forest. We assume the overall strength of the absorption can be captured by mean optical depth function $\tau_{\text{eff}}$ following Becker et al. (2013),

$$\tau_{\text{eff}}(z_{\text{abs}}) = 0.00958 \times (1 + z_{\text{abs}})^{2.90} - 0.132,$$  \hspace{1cm} (3)

where $z_{\text{abs}} = \lambda/1215.67$ Å − 1 is the redshift of the absorption systems. Apart from the mean optical depth, the series of stochastic Ly$\alpha$ forest are modeled as random Gaussian fluctuations $\omega(z_{\text{abs}}) \sim \mathcal{N}(0, \text{diag}(\sigma^2_{\text{abs}}))$. $\sigma_{\text{abs}}$ signifies the evolution of the absorption strength of the Ly$\alpha$ forest. Similar to Garnett et al. (2017) and Ho et al. (2020), we adopt the function form of $\sigma^2_{\text{abs}}(z_{\text{abs}})$ as

$$\sigma^2_{\text{abs}}(z_{\text{abs}}) = \omega_0 \phi (1 - \exp(-\tau_0 (1 + z_{\text{abs}})^\beta)) + c_0^2,$$  \hspace{1cm} (4)

where $\omega_0$, $\tau_0$, $\beta$, and $c_0$ are free parameters, and “$\phi$” denotes the element-wise product of two vectors.

Finally, the observational noise $\epsilon$ is modeled as $\mathcal{N}(0, \text{diag}(\sigma^2))$, $\sigma$ is the observed flux uncertainty. Putting them together, the observed quasar spectrum on the blue side can thus be written as

$$S_b = C_b \circ \exp(-\tau_{\text{eff}}(z_{\text{abs}})) + \omega(z_{\text{abs}}) + \epsilon_b.$$  \hspace{1cm} (5)

As for the red side, although there might a few minor absorption features, most of them are rejected in preprocessing, which leads to the simplified red-side model with only the continuum:

$$S_r = C_r + \epsilon_r,$$  \hspace{1cm} (6)

Finally, integrating the blue-side and red-side models leads to the final expression of the whole quasar spectrum

$$S = AC + \Omega + \epsilon = A\mu + AFh + A\Psi + \Omega + \epsilon,$$  \hspace{1cm} (7)

where

$$A = \text{diag}[\exp(-\tau_{\text{eff}}(z_{\text{abs}}))_{N_b}^{1, \ldots, N}],$$  \hspace{1cm} (8)

and

$$\Omega = [\omega(z_{\text{abs}})_{N_b}^{0, \ldots, N}]^T.$$  \hspace{1cm} (9)

We summarize all of the learnable model parameters in QFA and their corresponding dimensions in Table 1 and other variables in Table 2.

### 2.3. Model Training—Maximum Likelihood Estimation

Thus far, we have developed an analytic description of quasar spectra. We note that most variables written are stochastic variables, and the analytic formulae described should be treated as the mathematical operation (e.g., sum, product) on the random variables. The advantage of describing the quasar spectra as a stochastic process is that the probabilistic model defines the likelihood of any observed quasar spectrum. The best model would then be the one that maximizes the joint likelihood of all quasar spectrum observations. This section will derive the likelihood function and describe the model’s training via the maximal likelihood estimation objective.

#### 2.3.1. Log-likelihood

A particular advantage of QFA is that it is specifically designed in a way where most random variable components can be described as a high-dimensional Gaussian distribution. This naturally stems from the fact that the Ly$\alpha$ forest is assumed to be a Gaussian distribution in QFA. Although, in reality, the Ly$\alpha$ forest is not strictly Gaussian distributed (e.g., Bautista et al. 2015). We will return to this point in Section 5.7. However, this trade-off allows us to describe the likelihood in a compact Gaussian, which facilitates many other operations, such as marginalizing over nuisance parameters or masked pixels. We will show that the Gaussian assumption only incurs minor systematics in the continuum inference and other downstream tasks (see Section 4).

Recall that, for multivariate Gaussian distributions, the two properties below, which will come in handy in some of our derivations, hold (Johnson & Wichern 2007).

1. If $X$ distributed as $\mathcal{N}(\mu, \Sigma)$, then $BX + d$ distributed as $\mathcal{N}(B\mu + d, B\Sigma B^T)$.
2. For two independent multivariate Gaussian random variables $X \sim \mathcal{N}(\mu_X, \Sigma_X)$ and $Y \sim \mathcal{N}(\mu_Y, \Sigma_Y)$, their sum $X + Y$ is distributed as $\mathcal{N}(\mu_X + \mu_Y, \Sigma_X + \Sigma_Y)$.

Since $h$ is distributed as $\mathcal{N}(0, I)$ and $\Psi$ is distributed as $\mathcal{N}(0, \Sigma_{\Psi})$, it follows from the properties above that the quasar continuum $C$ is distributed as $\mathcal{N}(0, F F^T + \Sigma_{\Psi})$. Furthermore, the intrinsic quasar continua, Ly$\alpha$ forest absorption, and observational noise are independent. It follows trivially from Equation (7) and the two properties above that the distribution
of the quasar spectrum $S$ can be written as

$$ S \sim N(A\mu, A(FF^T + \Sigma\psi)A^T + \Sigma_\Omega + \Sigma_\epsilon), $$

where $\Sigma_\Omega = \text{diag}([\omega, \tau, \beta, c_0, F])$. The formula above characterizes the entire distribution of the quasar spectrum, given the model parameters $M = \{\mu, \sigma_\psi, \omega, \tau, \beta, c_0, F\}$. The training process of QFA is to find the best model parameters $M^*$, which maximizes the likelihood of all observed spectra.

In particular, for any observed quasar spectrum $(\lambda, S, z, \sigma_\epsilon)$, the log-likelihood of observing the spectrum is

$$ \mathcal{L}(S|\lambda, z, \sigma_\epsilon, M) = -\frac{1}{2}(N_{\text{pix}} \ln 2\pi + \text{Indet}\Sigma) $$

$$ + (S - A\mu)^T \Sigma^{-1}(S - A\mu), $$

where $\Sigma = A(FF^T + \Sigma\psi)A^T + \Sigma_\Omega + \Sigma_\epsilon$. And for a data set $\mathcal{D}$ with $N_{\text{spec}}$ independent observed quasar spectra $\{(X^{(i)}, S^{(i)}, z^{(i)}, \sigma_\epsilon^{(i)}), i = 1, 2, \ldots, N_{\text{spec}}\}$, their joint log-likelihood can thus be written as

$$ \mathcal{L}(\mathcal{D}|M) = \sum_{i=1}^{N_{\text{spec}}} \mathcal{L}(S^{(i)}|X^{(i)}, z^{(i)}, \sigma_\epsilon^{(i)}, M). $$

2.3.2. Model Regularization

In practice, we found that maximizing the likelihood in Equation (12) often leads to nonphysical local minima, i.e., continuum components with jagged features. Regularization tricks are necessary to facilitate the model convergence to local minima, better separating the distributions of the quasar continua and the Ly$\alpha$ forest.

We impose two regularization recipes. First, previous PCA-based works demonstrated that the principal components of quasar continua always fluctuate around zero. The results suggested that despite the great diversity of quasar continua, their variations are small compared to the mean continuum. This motivates us to assume a prior for all model parameters to be close to zero. In particular, we assign each parameter a Gaussian prior centered around zero, equivalent to what is known as the “L2 regularization” (Ng 2004). With the regularization term, the loss function reads

$$ L(M) = -\frac{1}{N_{\text{spec}}} \mathcal{L}(\mathcal{D}|M) + \alpha L_2(M), $$

where $\alpha$ is a hyperparameter that controls the regularization strength. A larger $\alpha$ gives heavier penalties to the weight of model parameters, leading to smaller model parameters. $L_2(M)$ denotes the L2 regularization, which is the square sum of all model parameters. We assume $\alpha = 0.1$ in this study.

Second, we also enforce that quasar continua are smooth profiles. To implement this regularization, we apply a running median filter along the wavelength direction to smooth each component in the factor loading matrix $F$ every 20 optimization epochs. In this study, we set the filter width to 31 pixels ($\sim$7–11 Å in the rest frame). Here, one optimization epoch is defined as the stochastic gradient-descent algorithm running over the entire data set.

2.3.3. Implementation

Traditionally, a factor analysis model is optimized through singular value decomposition or an expectation maximization algorithm (Barber 2012). However, both methods are difficult to be implemented for QFA because of the heteroskedasticity nature of QFA. Also, the elaborate modeling of QFA with $\sim 2 \times 10^4$ model parameters and the complex loss function (Equation (13)) calls for a better optimization algorithm.

We adopt the Adam optimization algorithm (Kingma & Ba 2014), a robust optimization method based on gradient descent, which has been widely used in deep learning, to find the best model parameters that optimize the likelihood of the observed data. We implement QFA via PyTorch (Paszke et al. 2019) to speed up the matrix operations with GPU resources. Due to the Gaussian nature of our model, the derivative of each parameter (see Appendix B) can be analytically derived, which we further harness to speed up the training process and optimize the GPU memory usage. For details, we refer the reader to Appendix A.

2.4. Continuum Inference

QFA depicts the joint distribution of the quasar spectra. The model breaks the degeneracy between the transmission fields and the continua by having two separate components for the continuum and the Ly$\alpha$ forest features. Consequently, given the best-fitted model $M^*$, one can obtain the posterior distribution of the quasar continuum $C$, conditioning on the observed quasar spectra, which we will elaborate on in this section.
QFA models the quasar continua $C$ as $C = \mu + Fh + \Psi$. We will neglect the “error” term $\Psi$ in the quasar continuum inference. The term $\Psi$ is designed to account for the stochastic residuals of quasar continua that cannot be explained by the linear model $\mu + Fh$. However, in practice, we found that $\Psi$ also incorporates other unaccounted for absorptions and observation noises and can lead to quasar continuum inference with jagged features. Thus, we neglect $\Psi$ when inferring the posterior distribution of the continua. Such treatment may result in poor continuum fitting at emission peaks in a few cases, and may also lead to an underestimation of the derived continuum fitting uncertainty. In Appendix 1, we provide a detailed explanation of the issue of underestimating the continuum fitting uncertainty and present a possible calibration process to counteract this shortcoming.

On top of that, since $F$ and $\mu$ are deterministic parameters, to derive the posterior of $C$, it suffices to evaluate the posterior of $h$. Recall that, given an observed spectrum and its properties $(\lambda, z, \sigma)$, the posterior of $h$ follows Bayes' rule:

$$P(h|\lambda, z, \sigma, M) \propto P(S|\lambda, h, z, \sigma, M)P(h),$$

where $h \sim N(0, I)$. Recall that, in Section 2.2, we derived that the conditional distribution of the quasar spectra $S$ can be written as

$$S_{\lambda,h,z,\sigma,M} \sim N(AFh + A\mu, (A\Sigma \phi A + \Sigma_\omega + \Sigma_e)^{-1}).$$

It follows from Equations (14) and (15) that, $h$ follows the distribution

$$h|\lambda, z, \sigma, M \sim N(\tilde{F}T\Sigma_e^{-1}\Delta, \Sigma_h),$$

where

$$\tilde{F} = AF$$
$$\Delta = S - A\mu$$
$$\Sigma_e = A\Sigma \phi A + \Sigma_\omega + \Sigma_e$$
$$\Sigma_h = (I + F^T\Sigma_e^{-1}F)^{-1}.$$

With the posterior distribution of $h$ at hand, the best-estimated continuum $C$ given the observed spectrum can be evaluated as the continuum fitting with the maximum posterior probability density, or

$$C = \mathbb{E}[Fh + \mu|\lambda, z, \sigma, M] = F\mathbb{E}[h|\lambda, z, \sigma, M] + \mu$$
$$= F\Sigma_h \tilde{F}T\Sigma_e^{-1}\Delta + \mu,$$

with the posterior variance

$$\text{Var}(C|\lambda, z, \sigma, M) = \text{diag}(F\Sigma_h F^T).$$

Modeling the full joint distribution of the quasar spectra comes with the advantage of taking the marginal distribution by integrating over all masked pixels. Various factors, such as limited wavelength coverage of the spectrograph, can render part of the quasar spectra unavailable. Furthermore, some strong absorbers might violate the assumption of our models where the absorptions are assumed to be Gaussian distributed; including them as the conditional information would, therefore, bias our inference.

Fortunately, since our posteriors are all Gaussian, taking the marginal distribution is trivial. More specifically, recall that for multivariate Gaussian distribution, the covariance matrix of the marginal distribution corresponds to the corresponding submatrices of the entire matrix, and the mean of the marginal distribution corresponds to the corresponding submean of the mean vector.

3. Data

In this study, we apply QFA to observation and mock data to demonstrate its capability compared to other widely adopted methods. We will apply QFA to observed quasar spectra from SDSS Data Release 16 (DR16; Lyke et al. 2020). Furthermore, we will also apply QFA for mock quasar spectra, of which we know their ground-truth continua, and evaluate their performance compared to existing methods.

3.1. SDSS DR16

The SDSS DR16 quasar catalog (DR16Q, Lyke et al. 2020) consists of a total of 750,414 quasar spectra conducted using the BOSS spectrographs on the 2.5 m wide-angle optical telescopes at Apache Point Observatory. The spectrographs cover a wavelength range from 3600 Å to 10400 Å at a spectral resolution of $\lambda/\Delta \lambda \approx$ 2000.

We exclude nonquasar spectra contaminants (e.g., star, blazar) by considering only spectra with a confident visual “quality flag.” CLASS_PERSON=3, as specified by the catalog. We compute the median $S/N$ of each quasar spectrum at rest-frame wavelength $1280 \pm 2$ Å and select only quasar spectra with median $S/N$ greater than 2.0 and redshifts $2 < z < 3.5$ to ensure relatively high data quality and reliable mean optical depth measurements. Note that as various redshift estimates may differ in DR16Q (Lyke et al. 2020), we use the $z$ column as a redshift reference. According to Lyke et al. (2020), the redshifts from the $z$ column are expected to be the least biased (although, arguably with a higher variance). Since QFA can generally deal with redshift variance, the training of QFA thus benefits from the least-biased estimator, which has led to our choice. We also discard those quasars flagged with broad absorption line systems (BAL). Although our algorithm can deal with masked pixels, e.g., the damped $Ly\alpha$ systems and missing data pixels (see Section 2.4), current techniques (Guo & Martini 2019) cannot yet reliably mask the BAL regions. In total, 90,678 quasar spectra met our criteria. The $S/N$ and redshift distribution of our final data set are shown in Figure 3. Also, our data selection criteria and their corresponding number of spectra after these successive cuts are shown in Table 3. We focus on the redshift interval $2 < z < 3.5$, as it encapsulates the majority of quasars used for cosmological studies, for example, BAO measurements (e.g., du Mas des Bourboux et al. 2020) and 1D $Ly\alpha$ forest power spectrum measurements (e.g., Chabanier et al. 2019).

Unlike some of the existing methods, our unsupervised learning algorithm not only harnesses information from high-$S/N$ spectra (e.g., $S/N \geq 10^2$ spectra in Pâris et al. 2011, and $S/N \geq 7$, $\sim 10^5$ spectra in Davies et al. 2018b), but also low-$S/N$ quasar spectra ($S/N \geq 2$ in this study) without the need for continuum labels. As shown in Figure 4, taking into account the low-$S/N$ quasar spectra enables us to cover a more extensive luminosity range than before (see Figure 2 in Lee et al. 2012), and admit a much larger training data set—of the order of $10^5$ as opposed to $\sim 10^2$–$10^4$ in the supervised learning.

5. https://data.sdss.org/sas/dr16/eboss/qso/DR16Q/

6. We calculate the monochromatic luminosity at rest-frame wavelength $1280 \pm 2$ Å.
states the number of quasar spectra left after these criteria are upheld. We specifically match the properties of the mock spectra to mimic the SDSS data to gauge the performance of QFA compared to other existing methods. Since the transmission field adopted to create the mock spectra truncates at \( z = 2.15 \), this leads to a minor difference in redshift coverage between the mock data set and the SDSS DR16 data set.

**Table 3**

| Criteria                                | Number of Spectra Left |
|-----------------------------------------|------------------------|
| All SDSS quasar spectra                | 750,414                |
| SDSS quality flag                      | 290,068                |
| S/N >2                                  | 220,960                |
| 2 < \( z < 3.5 \)                      | 125,156                |
| Not BAL                                 | 90,678                 |

Note.

\(^a\) We impose data selection criteria from top to bottom, and the right column states the number of quasar spectra left after these criteria are upheld.

methods (Suzuki et al. 2005; Pâris et al. 2011; Lee et al. 2012; Davies et al. 2018b; Reiman et al. 2020; Liu & Bordoloi 2021).

We transform the observed quasar spectra to the rest frame and renormalize them by dividing the whole spectra with the median value of the flux at the rest-frame wavelength 1280 Å. We perform piecewise sigma-clipping (3\( \sigma \)) to reject erroneous absorption lines and pixels and interpolate the flux to conform each spectrum onto a uniform rest-frame wavelength grid spanning from 1030–1600 Å with a logarithmic uniform spacing. For piecewise sigma-clipping, we adopt an adaptive filter window, adjusted based on the existence of the emission features: a 30 pixel window size (\( \sim 7–11 \) Å in rest frame) is used in the “smooth” region devoid of emission peaks, while a 10 pixel window size applies around emission peaks (\( \sim 2–4 \) Å in rest frame). The wavelength range covers the Ly\( \alpha \) forest regions but not the Ly\( \beta \) forest. As for quasar spectra whose rest-frame wavelength does not include the wavelength range of 1030–1600 Å, we consider those regions to be masked.

Finally, as we only model the Ly\( \alpha \) forest as in Section 2, the physical priors of our model do not include the damped Ly\( \alpha \) systems (DLAs). DLAs (Wolfe et al. 2005) are the population of strong absorbers with integrated neutral hydrogen (H I) column density \( N_{\text{HI}} \geq 2 \times 10^{20} \) cm\(^{-2} \), resulting in a broad absorption region (\( \Delta v \sim 10^3 \) km s\(^{-1} \)). The broad absorption features bias our inferences and ought to be masked. Fortunately, various techniques have been developed to give a robust estimation of the DLA regions (e.g., Garnett et al. 2017; Parks et al. 2018; Ho et al. 2020; Wang et al. 2022). In this work, we adopt the DLA classification results in the SDSS DR16 quasar catalog (Lyke et al. 2020), which include the central wavelength and column density information for each DLA. We mask the DLA regions within two times the equivalent width of each DLA (Draine 2011) and further perform damping wing correction as depicted in Lee et al. (2012). In Figure 5, we show an example of our data preprocessing. The absorption systems on the red side of the quasar spectrum are filtered by sigma-clipping (3\( \sigma \)). We also mask the DLA region, shown in blue.

### 3.2. Mock Spectra

Complementing our study of the SDSS data, we will also assume a set of mock quasar spectra of which the ground-truth continua are known. For the mock continua, we assume the PCA templates from Pâris et al. (2011). We adopt them as the basis quasar templates and fit our SDSS DR16 data set. The fits yield a set of quasar continua given by PCA continuum fitting. We then draw continua from the fitted data set.

As for the transmission fields, we adopt the mock transmission fields from the SDSS DR11 quasar-Ly\( \alpha \) forest mock data sets (Bautista et al. 2015). These mock transmission fields were well examined to mimic the actual non-Gaussian transmission fields and served as the baseline calibrator for the main BOSS Ly\( \alpha \) BAO measurements (Slosar et al. 2013; Font-Ribera et al. 2014; du Mas des Bourboux et al. 2020). To test that our methods can robustly deal with high column density absorbers, we further inject to the mock spectra the high column density absorbers, including damped Ly\( \alpha \) systems and Lyman limit systems in the SDSS DR11 quasar-Ly\( \alpha \) forest mock data sets.

As mentioned in Section 1, the PCA basis from Pâris et al. (2011) is constructed based on a limited number of (78, S/N \( \geq 10 \)) quasar spectra. However, in practice, a PCA basis may poorly represent the real-world continuum distribution, potentially yielding extremely poor performance for spectra outside the PCA space. To quantitatively assess each method’s robustness, we apply minor linear perturbations to the simulated quasar

---

\(^7\) http://www.sdss.org/dr12/algorithms/lyman-alpha-mocks
continua. More specifically, we assume a linear perturbation:

\[ C_{\text{perturb}}(\lambda) = C_{\text{PCA}}(\lambda) + M_0 + M_1 \left( \frac{\lambda / \text{Å} - 1030}{1600 - 1030} \right), \]  

(20)

where \( M_0 \) is drawn from a uniform distribution from \(-0.1\) to \(0.1\), \( M_1 \) is drawn from a uniform distribution from \(-0.2\) to \(0.2\), and \( C_{\text{PCA}} \) denotes the PCA constructed continua. The perturbation given by Equation (20) is no more than 10% at rest-frame wavelength 1280 Å. We re-normalize each continuum after the perturbation to ensure that all spectra have the same relative flux. Such a perturbed spectrum can be viewed as a weighted combination of the linear perturbation and the original spectrum. As the random perturbation is constrained to be sufficiently minor, the original and perturbed distributions should appear visually similar. Figure 6 demonstrates 100 unperturbed continua and 100 perturbed continua; perturbed continua have a similar distribution as the unperturbed continua. While the linear perturbation term we introduce may seem arbitrarily simplistic and unrepresentative of the actual continuum distribution, we employ it solely as a proof-of-concept experiment to evaluate each method’s robustness. As we show in Section 4.1, even minor perturbations cause existing supervised techniques to fail in generalizing, unlike QFA. Moreover, Section 3.1 indicates real quasar spectra likely demonstrate more complex variations than our linear perturbation. In that event, the advantage of QFA becomes more prominent.

Figure 4. The luminosity distribution of the SDSS DR16 data set. The bottom panel shows the dependency between the luminosity of quasars and the S/N, while the top panel gives the marginal distribution of quasar luminosity. The unsupervised nature of QFA enables us to incorporate low-S/N quasar spectra into the training set, leading to a more expanded luminosity coverage (\(10^{44}\) erg s\(^{-1}\) \(\sim\) \(10^{47.5}\) erg s\(^{-1}\)) in this study. The S/N > 7 selection criteria has been commonly adopted in previous works (e.g., Davies et al. 2018b).
Figure 5. Data preprocessing for the SDSS DR16 spectra. The quasar spectrum after preprocessing is shown in orange, and the original quasar spectrum from SDSS is in gray. The raw and preprocessed quasar spectra are normalized at rest-frame wavelength 1280 Å. We reject absorption lines in the red side and noisy pixels by sigma-clipping. The damped Ly\(\alpha\) systems are masked, as shown in the blue shaded region, and the damping wing beyond two times the equivalent width is subsequently corrected following the prescription in Lee et al. (2012).

Figure 6. Mock quasar continua adopted in this study. The unperturbed continua are generated based on the PCA method (Pâris et al. 2011), fitted with the SDSS DR16 data. The perturbed continua further assume a perturbation of 10% (see the text for details). The perturbation allows for a broader diversity of quasar continua beyond the PCA basis. The perturbed mock data serves as a more stringent test for the continua inference algorithm, testing their ability to generalize beyond the high-quality S/N spectra from which the PCA basis was derived.
Finally, we inject observational noise that mimics that from SDSS DR16 spectra. We divide our SDSS DR16 data set into five redshift bins from $z = 2$ to 3.5 with a redshift interval of 0.3. We randomly draw Gaussian random noise for each mock spectrum according to the SDSS DR16 observational uncertainty from the corresponding redshift bins. The S/N and redshift distribution of the mock spectra are shown in Figure 3. The S/N and redshift distribution of our mock spectra follow that from the SDSS DR16 data set; testing on these realistic mock quasar spectra thus allows us to demonstrate the validity of our statistical assumption introduced in Section 2 and yield a reliable quantification for the continuum recovery performance of different models.

4. Results

In the following, we will demonstrate the performance of QFA for quasar continuum inference, compared to two representative existing continuum fitting methods: the PCA continuum fitting method from Pâris et al. (2011) and PICCA continuum fitting from du Mas des Bourboux et al. (2020). It should be noted that there exist various variants of PCA-based techniques that differ in the size of the training data and the mapping methods between the red side and the blue side. However, most are not publicly available, and assessing them is beyond the scope of this study. Furthermore, as demonstrated by Bosman et al. (2021), their performance is comparable ($\lesssim 3\%$ difference), and they are subject to similar limitations as Pâris et al. (2011) when using a biased training set. Therefore, comparing against Pâris et al. (2011) shall not impact our conclusions significantly. We specifically compare these two representative methods (PCA and PICCA) because they best describe the state-of-the-art supervised learning methods (PCA) and unsupervised learning methods (PICCA) and have been widely demonstrated in practice. We leave discussion about other existing methods to Section 5.6. We note here that PCA only utilizes the red-side pixels to predict the blue-side pixels, whereas PICCA exclusively utilizes the blue-side pixels. In contrast, QFA encompasses both the blue-side and red-side pixels for prediction.

4.1. A Quantitative Assessment of QFA Compared to Other Existing Methods

First, we compare our model performance on the mock data set from which the ground-truth continua are known. Throughout this paper, we will consider the following metrics to quantify the performance of QFA. We define the absolute fractional error of continuum prediction as

$$\text{Absolute Fractional Error} = \left| \frac{C_{\text{pred}}(\lambda) - C_{\text{truth}}(\lambda)}{C_{\text{truth}}(\lambda)} \right| \times 100\%$$

(21)

to compare model performance at different wavelengths. We further quantify the overall performance by integrating over the wavelength. We define the absolute fractional flux error (AFFE) $|\delta C|$ as in Liu & Bordoloi (2021):

$$|\delta C| = \int_{\lambda_1}^{\lambda_2} \left| \frac{C_{\text{pred}}(\lambda) - C_{\text{truth}}(\lambda)}{C_{\text{truth}}(\lambda)} \right| d\lambda \int_{\lambda_1}^{\lambda_2} d\lambda \times 100\%$$

(22)

Briefly, AFFE measures, for individual spectrum, the average absolute error in the wavelength region of interest. Additionally, we define the absolute fractional uncertainty (AFFU) $|C_{\text{truth}}(\lambda)|$ as:

$$|\delta C| = \int_{\lambda_1}^{\lambda_2} \left| \frac{C_{\text{pred}}(\lambda) - C_{\text{truth}}(\lambda)}{C_{\text{truth}}(\lambda)} \right| d\lambda \int_{\lambda_1}^{\lambda_2} d\lambda \times 100\%$$

(23)

to better quantify the mean absolute bias of continuum fitting result. Compared with AFFE in Equation (22), AFFB performs as an extra metric to evaluate the difference between the ground-truth continua and fitted continua.

We also focus only on evaluating the performance at wavelengths bluer than the Ly$\alpha$ emission (for simplicity, “blue side”), i.e., from rest-frame wavelength 1040–1180 Å, which covers most parts of the Ly$\alpha$ forest but excludes the effect of the Ly$\beta$ forest (e.g., Yang et al. 2020) and proximity zone (e.g., Fan et al. 2006; Carilli et al. 2010). This is a conservative assessment for QFA, as both PCA and PICCA are mainly designed for recovering quasar continua in the blue side. The red-side continuum prediction for PCA is not optimized in practice. In contrast, PICCA does not support red-side continuum prediction. Although not shown here, QFA outperforms PCA on the red side of the quasar spectra (see Appendix D).

Figure 7 shows the absolute fractional error as a function of wavelength, evaluating the mock data set of 10,000 spectra. QFA yields more accurate and robust continuum predictions than PCA and PICCA on the region of interest in the blue side regardless of the perturbation. The quality of the predictions from all three methods shows some dependency with respect to the wavelength, but for different reasons. For PCA, which learns the blue-side continua from the red-side information, the larger relative error is because the correlation between the bluer pixels to the red continua is less prominent. We also see the same limitation, as shown in Figure 13. For PICCA, the fitted polynomial correction cannot fully describe the variations between the mean continuum and the continuum for individual spectrum, thus underfitting the continuum. The underfitting affects not only the bluer end but also the redder end.

For QFA, the reason for the more considerable uncertainty toward the bluer parts is two-fold. (a) The larger observational noise level in the bluer parts will inflate the prediction uncertainty, as shown in Figure 9. (b) As we fixed the mean optical depth function as physical prior (Equation (3)), the inconsistency between the predefined and ground-truth mean optical depth assumption may cause QFA to perform worse toward the bluer end, a limitation of our current model. We will return to this in Section 5.7.

Tables 4 and 5 summarize the overall performance, integrating over the wavelength. As shown, the continuum predictions from QFA are more accurate and remain robust even with perturbations on the mock continua. QFA achieves an accuracy of $\sim 2\%$ in AFFE in both mock data sets, outperforming PCA and PICCA. PCA performs the best for the 95th percentile of spectra. But we note that this is somewhat misleading because the mock continua are generated from the PCA template. Thus, it is unsurprising that PCA achieves the best performance in these limited cases when the PCA template perfectly matches the mock spectra. Nonetheless, even for the unperturbed case, QFA performs better generally. Importantly,8 A 4th-degree polynomial correction in PICCA is defined as $\sum_{i=0}^{4} a_i \lambda^i$, in which $a_0, \ldots, a_4$ are free parameters and $\lambda = \log \lambda$ is the wavelength in log space. We adopt first order in this study as in du Mas des Bourboux et al. (2020).
Figure 7. Continuum prediction error at different wavelengths. Shown are the median errors at individual wavelength pixels. The left figure shows the model performance on the unperturbed mock spectra, and the right figure shows the perturbed mock spectra. QFA yields an AFFE of -2% at 1040-1180 Å. PCA achieves similar prediction quality on the unperturbed spectra -3%. But for the perturbed data set, PCA performance degrades to -5%. PICCA is less susceptible to the perturbation, but its performance is worse than QFA, with an average AFFE of -3% in both data sets. Both PCA and QFA incur larger errors toward the bluer parts, while PICCA performs worse on both the redder and bluer ends.

Table 4

| Methodologies | Unperturbed | Perturbed |
|---------------|-------------|-----------|
| AFFE [%]      | 5th 50th 95th | 5th 50th 95th |
| QFA versus Truth | 1.29 2.47 4.82 | 1.17 2.52 5.02 |
| PCA versus Truth | 0.67 3.04 9.81 | 0.99 4.91 12.9 |
| PICCA versus Truth | 1.26 3.13 6.65 | 1.32 3.20 6.72 |

Note. We refer the reader to Section 3.2 for the definition of “unperturbed” and “perturbed” mock data sets. The results are evaluated over 10,000 mock spectra. The 5th, 50th, and 95th percentiles indicate the best, general, and worst performance, respectively, of the different methods on the 10,000 mock spectra. The bold values denote the best performance.

Table 5

| Methodologies | Unperturbed | Perturbed |
|---------------|-------------|-----------|
| AFFB (%)      | 25th 50th 75th | 25th 50th 75th |
| QFA versus Truth | -1.93 1.56 4.65 | -2.15 1.26 4.67 |
| PCA versus Truth | -8.25 0.29 7.35 | -10.4 1.06 11.6 |
| PICCA versus Truth | -2.80 1.38 5.90 | -3.05 1.38 5.95 |

Note. The bold values denote the best performance.

as we perturb the continua, the recovery of the PCA method degrades from -3% AFFE to -5% AFFE, but QFA remains resilient to the perturbation.

The unsupervised nature of PICCA makes it (similar to QFA) maintain similar performance on both data sets. Nonetheless, its predictions are worse than QFA because of its rigid parametric assumption, based on a first-order polynomial correction. PICCA attains a -3% in AFFE for both data sets. Among all three methods, QFA shows the smallest scatter, which is -1% AFFE for the best cases and -5% AFFE for the worst cases. PCA, in contrast, shows the largest scatter from <1% AFFE to >10% AFFE. Although PICCA only performs worse than QFA at the <1% level, its performance suffers from a larger scatter on a case-by-case basis, manifesting that the rigid polynomial correction performs subpar compared to QFA.

Finally, a desired continuum fitting algorithm should yield robust predictions regardless of the redshift or S/N of the quasar spectra. In Figure 8, we further evaluate the model performance as a function of redshift and S/N. Both PCA and QFA show little dependency on the redshift. PICCA shows a slightly stronger redshift dependence, which may be caused by the simultaneous fitting of both the mean optical depth function and quasar continua in PICCA as well as the redshift-dependent large-scale variance introduced in PICCA (see Equation (4) in du Mas des Bourboux et al. 2020). Since all three methods take into account the observational noise, their performance shows little to no evolution with S/N. As before, compared to PCA and PICCA, QFA gives the smallest scatter compared to the existing methods.

4.2. Application to SDSS Quasar Spectra

In addition to mock spectra, we apply QFA to the SDSS DR16 data set. We train a separate QFA model. Since we do not know the ground-truth continua for the SDSS spectra, we include 10,000 additional mock quasar spectra (without perturbation) in our training. These additional mock spectra, of which the ground-truth continua are known, are included to gauge the convergence of our models. We found that the median AFFE estimated on the mock data set in the Lyα region attains a precision of 2.36%, a performance on par with the case when the training set only consists of mock spectra. This demonstrates that an extension of the quasar distribution with the actual spectra does not adversely affect QFA. Therefore, we expect our model to achieve about the same accuracy in the SDSS DR16 data set as in the auxiliary mock data set in our training.

As for the actual SDSS spectra, since the ground truth is not directly accessible, we resort to evaluating the difference between the continuum estimates from QFA and those from PCA and PICCA. The differences are summarized in Table 6. As shown, there can be substantial deviations between the predictions from QFA and other methods. The predictions and PCA and QFA predictions can differ as much as >6% for -50% quasar spectra, and >10% for 25% quasar spectra, regardless of the S/N. Similarly, PICCA deviates >4% for
50% quasar spectra and \(\gtrsim5\%\) for 25% quasar spectra. These nonnegligible discrepancies beg the question of which methods infer the continua more accurately.

In Figure 9, we inspect the predictions from these different methods for SDSS spectra that show the most significant discrepancies. We show spectra that deviate at the 70th percentile level (of all of the SDSS spectra) and on the right, at the 90th percentile level. As shown in the figure, while the ground-truth continua are unknown, visual inspections suggest that the QFA estimates are more physically plausible. The inference from the PCA method tends to either overshoot or undershoot the observed spectra. PICCA, on the other hand, tends to diverge on the redder wavelength, which might be caused by the underfitting of the polynomial correction in PICCA.

These nonnegligible systematics of PCA and PICCA underline the importance of further advancing unsupervised continuum inference algorithms with QFA. As accurate continua are instrumental in constructing the Ly\(\alpha\) forest, these systematic differences might lead to errors in cosmology estimations via the Ly\(\alpha\) forest, which we will explore next.

### 4.3. Ly\(\alpha\) Forest Power Spectrum Measurements

The 1D power spectrum \(P_{\alpha}(k)\) of the Ly\(\alpha\) forest (e.g., Palanque-Delabrouille et al. 2013; Chabanier et al. 2019; Karaçaylı et al. 2020) is the telltale sign of the distribution of IGM in the distant Universe. The accurate quantification of the Ly\(\alpha\) forest power spectrum holds the key to many different sciences, including the thermal evolution of IGM (e.g., Gaikwad et al. 2021), the neutrino masses in our Universe (e.g., Rossi et al. 2017; Yèche et al. 2017), the dark radiation (e.g., Rossi et al. 2015), and the nature of dark matter (e.g., Iršič et al. 2017; Garzilli et al. 2019). The power spectrum of the Ly\(\alpha\) forest has much-renewed interest thanks to the ongoing and upcoming large-scale spectroscopic surveys, such as DESI (Schlegel et al. 2022), 4MOST (De Jong et al. 2019), and WEAVE (WEA 2016). However, any continuum residual, as demonstrated in the previous section, can cause a nonnegligible effect on the measurements of the power spectrum.

In the following, we will evaluate how well QFA can extract the Ly\(\alpha\) forest power spectrum as opposed to other existing methods. We will evaluate our performance, qualitatively on mock data sets described in Section 3.2. The transmission field of mock spectra translates into the ground-truth power spectrum, which can then be
compared with the recovered power spectrum from the different continuum extraction methods.

For a given spectrum with flux $S$, continuum $C$, and mean optical depth function $\tau_{\text{eff}}(z_{\text{abs}})$, the flux-transmission field can be evaluated as

$$ \delta = \frac{S}{C \exp(-\tau_{\text{eff}}(z_{\text{abs}}))} - 1. $$

To calculate the Ly$\alpha$ forest power spectrum, we then plug the flux-transmission field into the PICCA pipeline\(^9\) (du Mas des Bourboux et al. 2020). Briefly, the PICCA pipeline takes the flux-transmission fields as input and calculates the raw power spectrum of each transmission field through fast Fourier transform. The resolution effect, metal absorbers, observational noise, and other systematic errors from the data pipeline are taken into account to recover the underlying Ly$\alpha$ forest power spectrum. The final Ly$\alpha$ forest power spectrum is the ensemble average over those forest spectra in the corresponding redshift bin. We refer interested readers to Chabanier et al. (2019) for details of the Ly$\alpha$ forest power spectrum calculation. In practice, we calculate the transmission fields and, subsequently, the estimated power spectra with both the ground-truth continua and the estimated continua. The difference between the two tells us how much the imperfect continuum predictions from different algorithms imprint on the power spectrum estimate. In the following, we denote the ground-truth power spectrum as $P_{\text{truth}}(k)$, and the estimated one as $P(k)$.

\(^9\) https://github.com/igmhub/picca, version 4.2.0.

Figure 10 shows the measured Ly$\alpha$ forest power spectrum, comparing QFA with PCA and PICCA. We randomly select 5000 mock spectra from the perturbed mock data set and measure the 1D Ly$\alpha$ forest power spectrum over two redshift bins. At $2.3 < z_{\text{abs}} < 2.5$, QFA recovers the Ly$\alpha$ forest power spectrum to about $\sim 4\%$, while PCA incurs an error of $\sim 6\%$–$10\%$. At $2.9 < z_{\text{abs}} < 3.1$, QFA predicts a close-to-perfect Ly$\alpha$ forest power spectrum measurement, with an error of $\sim 1\%$, whereas PCA maintains an error of $\sim 5\%$. PICCA, as the state-of-the-art continuum fitting algorithm for the 1D Ly$\alpha$ power spectrum measurements (e.g., Chabanier et al. 2019), gives slightly worse performance ($\sim 4.72\%$ relative error) compared to QFA ($\sim 4.37\%$ relative error) at $2.4 < z_{\text{abs}} < 2.6$, and achieve comparable results ($\lesssim 1\%$ relative error) at $2.9 < z_{\text{abs}} < 3.1$. To thoroughly examine the performance of different models on the unperturbed data set, we present the 1D Ly$\alpha$ forest power spectrum measurements obtained using various continuum fitting methods in Appendix F. It should be noted that the increasing error in the 1D Ly$\alpha$ forest power spectrum toward smaller scales in Figure 10 is likely due to the bias in that specific redshift range, which arises from the intrinsic degeneracy between the mean optical depth function and quasar continua. We reported this bias in Section 4.1 and will further discuss the dependency of our model on the mean optical depth in Section 5.7. Additionally, we present an ablation study of the mean optical depth in Appendix H.2.

Figure 9. Model performance on the SDSS DR16 data set. We investigate the continuum predictions that show the most difference between PCA (top panels)/PICCA (right panels) and QFA. The left panels show the deviation at the 70th percentile (among the 37,548 SDSS spectra), and the right panels at the 90th percentile. PCA underestimates or overestimates the quasar continuum in these outlying cases, and continuum predictions given by PICCA tend to diverge toward both ends.

\(^\text{13}\) The Astrophysical Journal Supplement Series, 269:4 (30pp), 2023 November Sun, Ting, & Cai
2.3 relative deviation. The imperfect continuum reconstruction from the PCA leads to a larger systematic bias in the Lyα forest absorbers resides in at 1100–1180 Å, and the Lyα forest power spectrum measurement error with QFA is ≲ 1%, consistent with the continuum fitting results from 1100–1180 Å (see Figure 9). In contrast, PCA incurs an ≲ 5% fractional error. As discussed in Section 4.1, both the mean optical depth prior and the low S/N in the bluer regions might contribute to the QFA larger continuum prediction error at the bluer end.

For PICCA, as its average performance only differs from QFA at the level of 1% on the mock data set, it achieves almost the same precision as QFA for 1D Lyα forest power spectrum measurements, which is ≲ 4% at 2.4 < z < 2.6 and ≲ 1% at 2.9 < z < 3.1. Our results are consistent with Figure 7 in Chabanier et al. (2019), which found that PICCA introduces ~2%–4% relative error in the Lyα forest power spectrum measurements on the mock spectra. However, we caution that mock data sets might have simplified the question at hand. As shown in Table 6, PICCA and QFA still give noticeably different continuum predictions at the level of ≳ 3% for ~50% SDSS quasar spectra, and even ≳ 5% for ~25% SDSS quasar spectra, far larger than in the mock data sets, which is about ≲ 3% for ~75% mock quasar spectra (see Section 4.1). As such, a detailed comparison between QFA and PICCA on real-life data is needed to resolve this issue. But this is beyond the scope of this paper, and we will leave it to future work.

4.4. Quasar Outliers in SDSS

Massive data sets from modern-day large-scale spectroscopic surveys are bound to find unexpected interesting objects and demand systematic searches of such outliers. As QFA summarizes the ensemble of observed quasars into a probabilistic distribution, it provides a robust way to perform outlier detection. More specifically, given the QFA model, with the optimal model parameters $M^*$, the likelihood $L(S|\lambda, z, \sigma, M^*)$ of individual spectrum ($\lambda, S, z, \sigma$) can be evaluated according to Equation (11). The likelihood value indicates the probability of occurrence for each spectrum. A smaller likelihood value implies that the spectrum in question deviates from the majority and, hence, is an outlier.

As a proof of concept, we evaluate the likelihood (Equation (11)) for 37,548 quasar spectra without masked pixels from the SDSS DR16 data set. Although QFA can deal with missing pixels (see Section 2.4), we do not consider quasar spectra in the current outlier search because the marginalized likelihood has a different absolute scale than the likelihood with the full spectrum. We will defer the detailed investigations of outliers from the full SDSS catalog to future studies.

Figure 11 shows the density contours of the likelihood at different S/Ns. Note that quasar spectra with high S/Ns tend to have more concentrated probability density functions and hence a higher likelihood value. In comparison, quasar spectra with low S/Ns tend to have more dispersed probability density distribution functions and, therefore, a lower likelihood value. Therefore, a robust outlier search must consider the S/N difference between different quasar spectra. We apply the $k$-nearest neighbor outlier detection algorithm to PyOD (Zhao et al. 2019). The algorithm identifies outliers by sorting all data points according to the mean distance between each data point and its nearest $k$ neighbors. As such, each spectrum is only compared with its nearest $k$ neighbors with similar S/N and likelihood. We investigate the 99th percentile outliers under this metric, or 179 outliers from 37,548 quasar spectra.

As shown in the inset plots in Figure 11, further visual inspection confirms that the 179 outlier spectra show some unexpected spectral features. These outliers include (a)
undetected damped Ly$\alpha$ absorbers (DLAs); (b) associated damped Ly$\alpha$ absorbers (associated DLAs); (c) broad absorption lines (BALs); (d) Type II quasar feature—overly strong Ly$\alpha$ emission but weak continuum; (e) erroneous redshift estimation; and (f) misclassified nonquasar spectra. We provide more details of these outliers in Appendix E.

4.5. The Evolution of the Quasar Population

Recall that QFA decomposes quasar continua into a lower-dimensional embedding with a finite number of “factors” (see Section 2; we assume eight factors in this study). Importantly, unlike previous studies of PCA (Suzuki et al. 2005; Pâris et al. 2011; Davies et al. 2018b), the basis of which is derived from a limited number of high-quality quasar spectra, QFA can make use of all SDSS DR16 quasar spectra. This allows us to study, in detail, the evolution of the quasar population as a function of their luminosity and redshift, which we will explore in this section.

The latent factors in QFA are defined up to a rotation (see Section 2.1). To ensure that the basis is physically motivated, we adopt the varimax rotation (Kaiser 1958). Varimax determines the rotation by maximizing the sum of variances of the squared loadings. In other words, varimax seeks a decomposition that decorrelates the emission features and flat continuum. Compared to the orthogonal basis enforced by PCA-based methods, the nonorthogonal basis of QFA offers more flexibility, allowing for a more physically meaningful decomposition of quasar continua. As a result, varimax enhances model interpretability compared to Figure 7 of Pâris et al. (2011). As shown in the middle panel, the first component reflects the strength of the Ly$\alpha$ emission, the second panel recovers a power-law feature of quasar continua, and the third panel demonstrates the contribution from the CIV emission. We focus only on the three most notable factors and leave the other components to Appendix G. We note that our learned decomposition is consistent with those derived from modern PCA methods based on larger training samples. Both methods give a power-law component in addition to components with strong features corresponding to correlations between the strengths and shifts of a wide variety of broad emission lines (e.g., Davies et al. 2018b). However, QFA differs from PCA in two aspects: (a) QFA learns its components directly from millions of quasar spectra, whereas PCA components are derived from ad hoc quasar continua fitted by other automated algorithms; (b) the probabilistic description of QFA (Equation (2)) enables more flexibility, such as nonorthogonality, of the learned components.

Figure 12 further demonstrates the evolution of the latent factors ($h$ in Equation (2)) as a function of redshift (left panels) and luminosity (right panels). We evaluate the correlation with Pearson correlation coefficient $r$. The uncertainty of each correlation coefficient is estimated through Monte Carlo sampling, and we find $\lesssim 1\%$ uncertainties of the correlation coefficients for all eight factors. The left panels show that these three factors (as well as the other factors in Appendix G) do not exhibit any visible correlation with the redshift. The lack of dependency with redshift demonstrates that the quasar population has not evolved much from 1.83 Gyr ($z = 3.5$) to 3.33 Gyr ($z = 2$). In a corollary manner, even trained on moderate-redshift ($2 < z < 3.5$) quasar spectra, QFA might be able to infer high-redshift (e.g., $z > 5$) quasar continua robustly, which we will detail in Section 5.1.

The right panels illustrate that the factors contributing to the Ly$\alpha$ emission and the power-law factor do not correlate with the monochromatic luminosity of the quasar. Interestingly, the CIV

---

10 Varimax can also be applied to PCA components, but technically, it would no longer be PCA since the basis is not orthogonal.

11 Square loading denotes the element-wise product of the factor loading matrix $F$ in Equation (2).

---

Figure 11. Outlier detection with QFA. The background contour shows the likelihood evaluated with QFA of the 37,548 SDSS DR16 spectra. The colored symbols show outliers objects with low probability likelihood. We visually inspect these outliers and classify them into different classes, as shown in the legend. The inset displays some examples of these outliers (see also Appendix E). The black symbols within the contours are the values from 500 nonoutlier SDSS quasars for illustration.
emission factor is the only exception—it exhibits an unmistakable negative correlation ($r \sim -0.44 \pm 0.001$), as in fainter quasars tend to have a stronger CIV emission line and vice versa. This correlation is not unexpected and is consistent with what is known as the Baldwin effect (Baldwin 1977; Jensen et al. 2016). We note that, however, previous measurements (e.g., Jensen et al. 2016) focused only on the equivalent width of the CIV emission, and in our case, we study the factor embedding, which contains the CIV emission. Our result also demonstrates a slightly stronger negative correlation than the literature values (e.g., $\sim -0.35 \pm 0.004$ in Jensen et al. 2016), indicating that the latent embedding learned by QFA may better reflect the mechanism that produces the Baldwin effect.

5. Discussion

We proposed in this study an unsupervised statistical algorithm, QFA, for quasar continuum prediction. We demonstrated QFA reaches state-of-the-art continuum inference performance for $2 < z < 3.5$ quasar spectra, regardless of S/N, subsequently reducing the systematics in the Lyα power spectrum measurements by 2%–4% compared to the existing PCA methods. We also explored various downstream tasks with QFA, including outlier selection and the evolution of the quasar population. Below, we will further discuss some other prospects of QFA, putting it in the context of other existing methods. We will also dissect some of its current limitations as well as future prospects.

5.1. The Evolution of the Quasar Population and Its Implication to High-redshift Quasars

Although rare, high-redshift quasars remain uncontested probes to the study of the IGM, and the extended Lyα damping wing in high-redshift quasar spectra is still perhaps the best telltale sign of the neutral hydrogen fraction during the epoch of reionization. Despite their importance, studying high-redshift quasars also comes with unique challenges. As the IGM becomes primarily neutral at $z \gtrsim 6$, as shown in Figure 13, the Lyα forest obliterates the bluer flux in moderate-redshift quasar spectra, typically known as the Gunn–Peterson trough (e.g., Gunn & Peterson 1965; Fan et al. 2006). A key assumption for the study of high-redshift quasars (e.g., Davies et al. 2018b; Đuровčíková et al. 2020; Reiman et al. 2020) thus relies on the fact that we could extrapolate the quasar’s properties at lower redshift to their higher-redshift counterparts and determine the continuum at wavelengths bluer from the red-side information.

In this study, by analyzing the entire SDSS DR16 data set (see Section 3.1), we did not find any statistically significant evolution of the quasar evolution from $z = 2$ to $z = 3.5$ (Section 4.5), consistent with previous studies (e.g., Jensen et al. 2016). The lack of quasar evolution might lend credence to training on moderate-redshift quasars and applying them to high-redshift quasars. Complementary to this study, Yang et al. (2021) also reported no significant evolution from $z = 6.3$ to $z = 7.64$, except for a blueshift of the CIV emission line.

Assuming that we can extrapolate our inference to high redshift, as a precursor study, we have tentatively applied our QFA, trained on the 37,548 moderate-redshift SDSS quasars, and applied that to a $z = 5.42$ quasar J0231-0728 observed by the Keck Observatory (Rafelski et al. 2012, 2014). Since the saturated transmission fields violate our assumption on the Lyα forest (see Section 2.2), we consider all of these wavelength pixels bluer than the Lyα emission to be masked during the inference. To adjust for the difference in resolution between the spectra obtained from Keck and SDSS, we downsample the high-resolution Keck spectra ($R \approx 7000–10,000$;
Rafelski et al. 2012, 2014 to match the SDSS resolution ($R \approx 2000$; Lyke et al. 2020). We then carry out the same preprocessing procedures as elaborated in Section 3.1. As shown in Figure 13, QFA yields a visually plausible quasar continuum. Interestingly, unlike the other inference shown at low redshift (Figure 9), as we deprive the blue information from the QFA, the inference uncertainty grows significantly toward the bluer end. This is not unexpected because there is a decline in correlation between the blue-side quasar continuum and the red-side quasar continuum. And hence when inferring only the blue side from the red side, like previously done in the PCA-based method, the inference uncertainty increases. We note that, as a preliminary experiment, the training sample and inference process are not optimized for high-redshift quasar spectra; compared to previous high-redshift quasar continuum prediction works (e.g., Davies et al. 2018b), QFA’s main potential is with increasing the diversity of the training samples and being probabilistic; more detailed considerations will be addressed in future works.

5.2. Probabilistic Inferences and Impact for Cosmological Measurements

In this study, we showed that the accuracy of continuum fitting of QFA can further suppress to the systematic bias of the Ly$\alpha$ power spectrum measurements to 1%–3% depending on the redshift (Section 4.3). But other than the accurate recovery, perhaps an even more critical innovation of QFA is that it also provides the posterior of the continuum. Thus far, most existing continuum fitting methods only provide a deterministic measurement of the continuum (e.g., Suzuki et al. 2005; Pâris et al. 2011; Liu & Bordoloi 2021). The classical approach is that, when inferring the cosmology, the uncertainty introduced from the continuum inference is calibrated through synthetic data (e.g., Chabanier et al. 2019). However, this approach comes with the danger of biased estimates, especially at the percent level. For example, as we have also shown in this study (see Sections 3.2 and 4.1), any PCA-generated mocks might not capture the full diversity of quasars and may lead to a biased calibration. We note that modern PCA methods (e.g., Davies et al. 2018b; Bosman et al. 2021) estimate the continuum fitting uncertainty directly from the training and test spectra.

QFA does not rely on such post hoc calibration, and it also provides also the posterior of the continuum prediction. The sampling of the posterior continuum is analytic and straightforward. In practice, once the posterior distribution of the latent factor ($h$) is computed (Equation (16)), we can then sample the posterior distribution of the latent factor and subsequently the corresponding posterior quasar spectra (Equation (2)). The probabilistic nature of QFA might prove important for future missions because the posterior can be directly integrated into cosmological measurement pipelines (e.g., du Mas des Bourboux et al. 2020), leading to a more ab initio Bayesian uncertainty quantification for cosmological parameters (e.g., Eilers et al. 2017; Simon et al. 2022; Gerardi et al. 2023).

5.3. Dissecting the Physics behind the Quasar Continua

Since QFA assumes a latent factor decomposition of the quasar continua, it projects the high-dimensional quasar continua into low-dimensional latent embedding (in our case, eight latent factors; see Figures 12 and 21). Compared to PCA methods, latent factor analysis is not confined to an orthogonal basis. As shown in Section 4.5, this flexibility in choosing the basis has led to a somewhat more sensible decomposition of the
quasar continua. In particular, as shown in Figure 12 and Appendix G, most of the components consist of a handful of prominent broad features.

It has long been postulated that the broad emission lines in quasar spectra are produced by the line-emitting gas in the broad-line regions and are closely associated with the accretion process of supermassive black holes (e.g., Shen et al. 2011; Yang et al. 2021). Recall that the properties of black holes (primarily, the mass) determine the temperature and pressure profile of the accretion disk. As such, the various latent components learned by QFA, e.g., the component representing the CIV emission line, might thus reveal the physics of the accretion disk, including the virial motion of the line-emitting gas in the broad-line region (e.g., Czerny & Hryniewicz 2011), and the radiation-driven outflows (e.g., Meyer et al. 2019).

Since the spectral embeddings from QFA likely relate to supermassive black hole properties, an intriguing possibility would be to find the mapping between supermassive black hole properties (including their masses and Eddington ratios) and this spectral embedding (see Eilers et al. 2022). In the same vein, a possible way to further improve on QFA is to harness the subset of quasar spectra for which we know the more “ground-truth” properties of supermassive black holes (e.g., through reverberation mapping; Vestergaard & Peterson 2006). Guided by this small set of “labels,” an even better latent factor decomposition might be possible through a mixture of supervised and unsupervised training.

Finally, QFA might also help us study changing-look AGNs —AGNs that show substantial time variations due to the changes in the accretion process of the supermassive black holes, outflows, and clouds (e.g., Ricci & Trakhtenbrot 2022). While the signatures in the spectral space might be subtle, the variation in the lower-dimensional embedding learned by QFA should be more prominent.

5.4. The Correlation and the Information Content in Quasar Continua

The generative nature of QFA also allows us to sample continua and evaluate pixel-wise correlations, as illustrated in Figure 14. The correlation matrix was estimated based on quasar continua sampled from a well-trained QFA model. By doing so, we reveal the information being leveraged. As a whole, the correlation matrix QFA recovers is reminiscent of that found in other studies (e.g., Suzuki et al. 2005; Páris et al. 2011). The blue–red intercorrelation has a typical value of 0.1–0.5, weaker than the intracorrelation (0.2–0.9). The blue–red intercorrelation being weaker than the intracorrelation has motivated this study. Recall that most existing continuum fitting methods (e.g., Suzuki et al. 2005; Davies et al. 2018b; Dušovčíková et al. 2020; Reiman et al. 2020; Liu & Bordoloi 2021) rely on relating the blue side with the red side. QFA goes beyond this and also considers the entire quasar spectrum when making the continuum inference. The fact that QFA also harnesses the information in the blue might explain the superior performance of QFA.

Comparing Figures 9 and 13 further supports this idea. As shown in Figure 9, the continuum posterior remains tightly constrained for the moderate-redshift spectra, where we can model the transmission field on the blue side and condition them. However, the continuum uncertainty inflates at the bluer end for the high-redshift quasar when we can only harness the information from the red (Section 5.1). The significant deviation between QFA and PCA methods when applied to the SDSS spectra also resides in the blue, further suggesting that the PCA method fails to capture the blue continuum accurately because the blue–red correlation subsides.

Interestingly, compared to the literature studies (e.g., Figure 2 in Páris et al. 2011), QFA favors a weaker blue–red intercorrelation. While QFA has a correlation of 0.1–0.5 between the Lyα forest region (1030–1216 Å) and the Lyα-CIV region (1216–1550 Å), previous studies have attained a correlation value of 0.3–0.5. Similarly, between the blue side and the region redder than the CIV emission line (λ_{RF} > 1550 Å), QFA suggests a correlation value of −0.4 to −0.1, whereas the literature values cluster around −0.6 to −0.4.

This difference might suggest that when training only on the handful of (~100) high-S/N samples, previous methods might have overestimated the correlation, leading to poorer generalization ability (also borne out with the SDSS test in this study). When QFA is trained on ~200 S/N ≥7 SDSS DR16 quasar spectra, we recover correlation values closer to these literature values (~0.3–0.9 intracorrelation between the Lyα forest regions and the regions between the Lyα and the CIV emission; ∼−0.8 to −0.4 intracorrelation between the Lyα forest regions and the regions with wavelengths longer than the CIV emission line). However, when we expand the sample size to learn the quasar properties from the diverse range of quasars in SDSS, the intracorrelation diminishes. We conducted experiments to validate whether the decrease in correlation was caused by the low S/N in our sample. However, we observed a similar decrease in correlation even when using only high-S/N (e.g., S/N ≥7) spectra for training, suggesting that including a low-resolution S/N sample in the training has little effect on the estimated correlation. Thus, our results suggest that the previously observed intracorrelation might be overly optimistic, which further underlines the importance of harnessing the information from the blue side when reconstructing the quasar continua.

5.5. The Power of Unsupervised Learning

In addition to harnessing the information in the blue, a key advantage of QFA is that it directly learns the distribution of quasar continua and the transmission fields by modeling the entire set of observed spectra through their combined effects. Most existing methods thus far rely on supervised learning (e.g., Suzuki et al. 2005; Páris et al. 2011; Davies et al. 2018b; Dušovčíková et al. 2020; Reiman et al. 2020; Liu & Bordoloi 2021), learning the mapping between red-side continua to blue-side continua. The key innovation of our study is that QFA can learn the continua directly from the observed spectra, alleviating the need to have a training set with ground-truth continua. Unsupervised learning aims to learn the distribution of quasar spectra through appropriate prior knowledge of the underlying data structure (Section 2). In the case of QFA, the prior knowledge we impose comes from the underlying structure of the continua and the transmission fields.

While we made a clear distinction between supervised learning and unsupervised learning to highlight the fundamentally different concept that underscores QFA, in many ways the two methods are related. Supervised and unsupervised learning reflect our different prior beliefs of the system. In the case of supervised learning, the model emphasizes the validity of the
training quasar continua. However, the lack of high-quality ground-truth continua can often lead to a biased model. In the unsupervised learning of QFA, we relay a different form of prior knowledge, focusing only on our understanding of how the quasar continua and transmission fields operate by assigning a specific functional form for these individual components. As we have seen in this study, this weaker physical prior, bolstered by the massive data sets we have garnered, can lead to a far more superior performance in continuum inference.

Finally, while we focus on unsupervised learning in this study, a hybrid form of weak unsupervised learning, fine-tuned with a subset of “supervised labels,” has led to many new ideas in the machine-learning community. These methods have coined the term “semisupervision” (e.g., Berthelot et al. 2019) or “self-supervision” (e.g., Chen et al. 2020; Xie et al. 2021). The same concepts have also seen some successes in their application in astronomy, including galaxy morphology classification (Walmsey et al. 2022), modeling stellar spectroscopy (O’Briaín et al. 2021), and transient identifications (e.g., Villar et al. 2020; Marianer et al. 2021; Slijepcevic & Scaife 2021). The future of quasar continuum inference might therefore lie in such a hybrid mode, comprehensively using all available ground-truth labels when they are available and at the same time, harnessing our insight into the underlying physical process of quasars, as we did in this study.

### 5.6. Other Existing Methods

In this study, we compare QFA with only two methods, the PCA algorithm proposed by Pâris et al. (2011) and PICCA. We focus on only these two methods because they remain some of the most adopted methods in cosmological measurements (e.g., Font-Ribera et al. 2014; Chabanier et al. 2019; du Mas des Bourboux et al. 2020; Bosman et al. 2021). But we note that there are many other more advanced PCA-related methods (e.g., Lee et al. 2012; Davies et al. 2018b; Šurovčíková et al. 2020) being proposed since the seminal paper of Pâris et al. (2011).

While these variations have undoubtedly enriched the possibilities of the PCA-based methods, Bosman et al. (2021) performed a comprehensive study of the differences between...
these methods and concluded that the lack of a ground-truth training set bottlenecks the PCA-based methods, rendering them to offer a similar performance (within \( \sim 10\% \)) for the continuum prediction. As different PCA-based methods demonstrate similar performance, we contend that adopting these different variances is less likely to alter the qualitative conclusions of this study. However, the limitations of these PCA-based methods, as demonstrated in Sections 4.1 and 4.2, are that they rely primarily on predefined continua that invariably lead to biased training sets, as discussed in Section 3.1, and only utilize information from the red side. These shortcomings have not been addressed in updated methods. Nonetheless, we note that modern PCA-based methods (e.g., Davies et al. 2018b) were primarily developed for high-redshift quasar spectra in which little blue-side information is available. As such, as discussed in Section 5.1, the advantages of QFA compared to PCA methods are likely to be less prominent for high-redshift quasar spectra, apart from enlarging the training data set to include fainter and lower-S/N quasars and being probabilistic.

Finally, in recent years, the study of quasars has also seen the rise of deep-learning methods (e.g., Liu & Bordoloi 2021). Comparing all of these methods is clearly beyond the scope of this study. However, we note that most of these deep-learning methods still focus on supervised learning and thus inherit the same problem as the PCA-based approaches.

5.7. Caveats and Limitations

We demonstrate that the unsupervised nature of QFA has led to superior performance in continuum inference. However, the model assumptions also currently limit the performance of QFA. In particular, we assume a predefined (not trainable) mean optical depth because the mean absorption degenerates with the quasar continua. While prior studies (e.g., Faucher-Giguère et al. 2008; Becker et al. 2013) have largely agreed on mean optical depth measurements, at least at the redshifts of interest in this study, the mean optical depth is not well constrained at higher redshifts. As we expand beyond the current redshift range, the data may necessitate a model in which the mean optical depth is trainable. We also discuss the effects of different mean optical depth functions in Appendix H.2. We found that the variations in continuum predictions are commensurate with mean optical depth function measurements. This suggests the importance of evaluating alternative mean optical depth functions in practice.

Perhaps the more important limitation of QFA is the assumption that the \( \text{Ly} \alpha \) forest constitutes an independent Gaussian distribution. In QFA, we assume such a distribution because the independent \( \text{Ly} \alpha \) forest assumption is essential for the analytic derivation. While this assumption is adequate for this study (because, for SDSS quasar spectra with typical resolution \( \lambda/\Delta \lambda \approx 2000 \), the correlations between adjacent pixels are generally weak), the assumption is clearly false in detail. For instance, significant absorbers in the IGM, such as DLA, have demonstrated the absorption in the adjacent wavelength pixel is anything but uncorrelated. On top of that, Farr et al. (2020) showed that the \( \text{Ly} \alpha \) forest also contains higher-order moment information beyond the Gaussian assumption.

Furthermore, the stochastic “error” term \( \Psi \), which stands for the differences between the quasar continuum and its dimension-reduced form \( \mu + F_h \), is assumed to be independent over wavelength, which is undoubtedly an oversimplified assumption since coherent structures can contribute to the continuum fitting error. However, in order for the stochastic optimization process to converge, we deem this simplification necessary and justified as \( \Psi \) is typically small compared to \( \mu + F_h \), i.e., \( \sim 3\% \) from our experiments.

Due to these limitations, this is why, when deriving the \( \text{Ly} \alpha \) power spectrum in this study, we only use QFA to make the inference on the continuum instead of directly using the inferred transmission field. A better QFA thus requires us to develop a more generalized formalism that can take into account the interpixel correlation and the higher-order moment simultaneously while ensuring that the models are still analytic or easily optimized. This is undoubtedly a tall order that we will leave to future studies.

6. Conclusion

In this study, we propose an unsupervised learning method, QFA, to infer quasar continua. QFA learns the distribution of quasar continua and the transmission field directly from the ensemble of observed spectra, regardless of their S/N. QFA does not depend on any predefined continua as a training set and can provide uncertainty quantification of the continua. The probabilistic nature of QFA allows the method to deal with missing pixels, capture a more physically motivated lower-dimensional embedding, and find spectral outliers. Our main findings are summarized as follows:

1. Testing on mock data sets, we demonstrate that QFA reaches state-of-the-art performance, \( \sim 2\% \) absolute fractional flux error at wavelength bluer than the \( \text{Ly} \alpha \) emission and \( \lesssim 1\% \) at wavelength redder the \( \text{Ly} \alpha \) emission, as opposed to the \( \gtrsim 3\% \) error from the PCA-based method and PICCA.

2. In addition to a better mean recovery, QFA also incurs the least case-by-case scatter. The absolute fractional flux error from QFA in the continuum recovery ranges from \( \sim 1\% \) in the best cases to \( \sim 5\% \) in the worst cases. In contrast, PCA has an error of \( 1\%–10\% \), and PICCA \( 1\%–7\% \).

3. QFA generalizes better. When introducing a 10% linear perturbation beyond the training set, the errors from the PCA-based method double, while QFA remains adaptable and achieves the same performance.

4. We further applied the method to the SDSS DR16 data set and showed that QFA yields more robust continua. The PCA-based method can deviate \( \gtrsim 6\% \) from the QFA’s inferences for 50% of the SDSS spectra, and \( \gtrsim 10\% \) for 25% of the SDSS quasar spectra, while PICCA gives \( \gtrsim 3\% \) differences for 50% SDSS quasar spectra from QFA’s inferences and \( \gtrsim 5\% \) for 50% SDSS quasar spectra.

5. QFA’s superior performance in extracting the continuum reduces the bias when estimating the \( \text{Ly} \alpha \) forest estimation to \( 3\%–4\% \) at \( z = 2.4–2.6 \) and to \( \lesssim 1\% \) at \( z = 2.9–3.1 \), which is a slight improvement to PICCA. In contrast, PCA induces a systematic bias of \( 5\%–7\% \) for both redshift ranges.

6. Through the latent embedding extracted through QFA, we illustrate that the quasar population does not manifest any detectable evolution from \( z = 2 \) to 3.5.
In a nutshell, QFA provides a new framework to model the quasar populations. We showed that, even with a minimal prior assumption on the quasar continua and the transmission, modern-day optimization frameworks allow us to model the quasar continua and transmission field from a vast set of quasar spectra simultaneously, without ad hoc human intervention. As large-scale spectroscopic surveys such as DESI and 4MOST will continue to revolutionize the large-scale spectroscopic surveys such as DESI and 4MOST quasar populations. We showed that, even with a minimal prior history of cosmic evolution. To maximally extract every bit of information from all of these heterogeneous quasar data sets, allowing the quasars to transmit to us the information about their own formation as well as the history of cosmic evolution.

7. Code and Data Availability

To ensure reproducibility, all source codes of Quasar Factor Analysis are made publicly available on Zenodo (Sun et al. 2023; doi:10.5281/zenodo.8050660). The data used in this study can be found on Zenodo at doi:10.5281/zenodo.8050660.

Acknowledgments

The authors thank Paulo Montero-Camacho and Pablo Renard for the detailed comments on the manuscript. The authors are also appreciative of the useful comments and discussions from Song Huang, Bradley Greig, Siwei Zou, Yunjing Wu, Xiaojing Lin, Ben Wang, Jiaqi Zou, Mingyu Li, Youquan Fu, Yongming Liang, Rongmon Bordoloi, Bin Liu, and Jiashu Pan. Z.S. particularly acknowledges Mingsheng Youquan Fu, Yongming Liang, Rongmon Bordoloi, Bin Liu, and Jiashu Pan. Z.C. are supported by the National Key R&D Program of China (grant No. 2020YFA0404503), the National Science Foundation of China (grant No. 12073014), and the science research grants from the China Manned Space Project with No. CMS-CSST2021-A05.

Appendix A

Complexity Analysis

The training and inference process of QFA involves manipulations of large matrices, which can be computationally expensive. Furthermore, direct large matrix inversion can lead to catastrophic divergence problems in practice. A crucial part of QFA is thus to design robust computation strategies, which we will detail in this Appendix.

The computational complexity for matrix inversion and log determinant are both $O(N_{\text{pix}}^3)$. The dimension of quasar spectrum ($N_{\text{pix}}$) and the number of quasar spectra is large ($N_{\text{spec}}$). As such, naive implementations of evaluating the loss function in Equation (11) and inhering the posterior distribution of quasar continua in Equation (16) can lead to a prohibitively large computational complexity. However, we can make use of the special structure of the covariance matrices of QFA to reduce the computation (Garnett et al. 2017; Ho et al. 2020).

In particular, the covariance matrix $\Sigma$ given by QFA has the form as $MM^T + D$, in which $\Sigma$ is an $N_{\text{pix}} \times N_{\text{pix}}$ matrix, $M$ is an $N_{\text{pix}} \times N_h$ matrix, and $D$ is an $N_{\text{pix}} \times N_{\text{pix}}$ diagonal matrix. Inverting this large matrix is computationally expensive, even when it is possible. Fortunately, we can reduce the complexity of matrix inversion by applying the Woodbury identity:

$$
(MM^T + D)^{-1} = D^{-1} - D^{-1}M(I + M^TD^{-1}M)^{-1}M^TD^{-1},
$$

where $I$ is an $N_h \times N_h$ identity matrix. Thus, we can compute the inverse matrix of $\Sigma$ through calculating the inverse of a much smaller $N_h \times N_h$ matrix $I + M^TD^{-1}M$, which is significantly more numerically stable. This simplification leads to a total computational complexity of only $O(N_{\text{pix}}^2 N_h^3)$.

Similarly, for the determinant, we apply the Sylvester determinant theorem,

$$\log\det(MM^T + D) = \log\det D + \log\det(I + M^TD^{-1}M),$$

reducing the complexity of computing the log determinant from $O(N_{\text{pix}}^3)$ to $O(N_{\text{pix}}^2 + N_h^3)$.

Appendix B

Derivative of the Loss Function

When optimizing the model parameters to maximize the likelihood (Equation (12)) of the data, derivatives of the loss function (here, the likelihood function) with respect to the model parameters are needed to update the model parameters during the optimization process. Although the derivatives (or “gradients”) can be calculated via auto-diff algorithms in deep-learning frameworks, implementing the analytical expressions of the derivatives when it is possible can speed up the GPU calculations.

Recall that, for a single spectrum ($\lambda$, $S$, $z$, $\sigma_z$), its likelihood function $L(\lambda, S, z, \sigma_z | M)$ is shown in Equation (11). For ease of derivation, we further rewrite $\Sigma$ and $\Delta$ as

$$\Sigma = AFF^TA + A\Sigma_{\phi}A + \Sigma_{\sigma} + \Sigma_v$$
$$\Delta = S - A\mu$$
$$\tilde{A} = \text{diag}^{-1}(A)$$
$$\varphi = \text{diag}^{-1}(\Sigma_v)$$
$$h(z_{\text{abs}}) = (1 - \exp(-\tau_0(1 + z_{\text{abs}})^3) + c_0)$$
$$f(z_{\text{abs}}) = h(z_{\text{abs}})^2.$$  

Here we define “\text{diag}” as the operation of returning a matrix whose diagonal elements are the input vector, and “\text{diag}^{-1}” denotes the operation of returning the diagonal elements of the input matrix. “\text{Sum}” denotes the summing up of all elements within a vector or a matrix. It is easy to check the derivatives of Equation (11) for all of the model parameters are as follows:
These analytic expressions are explicitly calculated and included in the gradient-descent process to optimize for the model parameters.

**Appendix C**

**Examples of Continuum Reconstruction on the Mock Data Set**

To help readers intuitively evaluate the results of continuum fitting presented in Section 4.2, we provide eight examples in which PCA/PICCA fails but QFA performs well. These examples are taken from the mock data sets discussed in Section 3.2, and therefore, the ground-truth continua are shown. From Figures 15 and 16, it is evident that inferring ground-truth continua from moderate-resolution quasar spectra is a challenging task, even for high-S/N quasar spectra.

\[
\frac{\partial \mathcal{L}}{\partial \sigma^2_{\phi}}(\lambda, S, z, \sigma|\mathcal{M}) = \frac{1}{2} \tilde{A} \text{diag}^{-1}(\Sigma^{-1} - \Sigma^{-1} \Delta \Sigma^{-1}) \circ \tilde{A}
\]
\[
\frac{\partial \mathcal{L}}{\partial \sigma_{\omega}}(\lambda, S, z, \sigma|\mathcal{M}) = \frac{1}{2} \text{diag}^{-1}(\Sigma^{-1} - \Sigma^{-1} \Delta \Sigma^{-1}) \circ f(z_{\text{abs}})
\]
\[
\frac{\partial \mathcal{L}}{\partial \eta_{0}}(\lambda, S, z, \sigma|\mathcal{M}) = -\text{Sum} \{ \text{diag}^{-1}(\Sigma^{-1} - \Sigma^{-1} \Delta \Sigma^{-1}) \circ \tilde{\omega} \circ f(z_{\text{abs}}) \circ h(z_{\text{abs}}) \circ (1 + z_{\text{abs}})^{2} \}
\]
\[
\frac{\partial \mathcal{L}}{\partial \beta}(\lambda, S, z, \sigma|\mathcal{M}) = -\text{Sum} \{ \text{diag}^{-1}(\Sigma^{-1} - \Sigma^{-1} \Delta \Sigma^{-1}) \circ \tilde{\omega} \circ f(z_{\text{abs}}) \circ h(z_{\text{abs}}) \circ \eta_{0} (1 + z_{\text{abs}}) \circ \ln (1 + z_{\text{abs}}) \}
\]
\[
\frac{\partial \mathcal{L}}{\partial \omega_{0}}(\lambda, S, z, \sigma|\mathcal{M}) = -\text{Sum} \{ \text{diag}^{-1}(\Sigma^{-1} - \Sigma^{-1} \Delta \Sigma^{-1}) \circ \tilde{\omega} \circ f(z_{\text{abs}}) \circ h(z_{\text{abs}}) \circ (1 + z_{\text{abs}})^{2} \}
\]

(28)

Figure 15. Similar to Figure 9. We show examples on the mock data set in which PCA fails but QFA performs well.
Appendix D
Continuum Reconstruction on Wavelengths Longer than \Ly\alpha Emission

In Section 4, we focused on the model performance at the \Ly\alpha forest regions because the \Ly\alpha forest regions carry valuable information on the density fluctuations. For completeness, in this Appendix, we also evaluate the model performance on wavelengths longer than the \Ly\alpha emission (rest-frame wavelength from 1280–1600 Å). The results are summarized in Table 7.

QFA yields, on average, more accurate continuum predictions on the red side ($\lesssim$1% AFFE) than the blue side ($\sim$2% AFFE). This is to be expected because the model can focus on only capturing the continuum without being set back by degeneracy between the quasar continua and the \Ly\alpha forest. As before, the unsupervised nature of QFA enables it to work robustly for both the perturbed and unperturbed mock quasar spectra. QFA reaches $\lesssim$1% AFFE on the red side for both mock data sets. As shown in Figure 17, QFA performs well across the entire wavelength range, achieving an accuracy of $\lesssim$1% over the full wavelength coverage. In contrast, similar to its performance in the \Ly\alpha forest regions (see Section 4.1), PCA is less robust compared to QFA due to its supervised nature. It yields $\sim$1% AFFE in the unperturbed data set but...
∼4% AFFE in the perturbed data set. Also, PCA’s performance shows a larger scatter than QFA in both data sets.

We also apply the methods to the SDSS DR16 data set and illustrate two randomly selected examples in Figure 18. The figure demonstrates that QFA also works well on the observed spectra. QFA gives reasonable continuum predictions in both examples: the QFA predictions generally pass through the mean flux and are not affected by the absorptions on the red side. PCA, in contrast, yields subpar continua, generally overshooting or undershooting the observed spectra.

Appendix E
Examples of Quasar Spectra Outliers in SDSS DR16

In Section 4.4, we have discussed how QFA can effectively find spectral outliers. From the 37,548 spectra we studied in SDSS, we determine 179 outliers (see Section 4.4 for details). We visually inspected these outliers and classified them based on their features. Figure 19 shows a representative spectrum from each of these visual classes.

These classes are defined, from top to bottom, primarily based on the following features: (1) Quasar spectra with unusually strong Lyα emission. The continuum is almost 20 times weaker than the Lyα emission in the example shown. (2) Quasar spectra with associated DLA (e.g., Finley et al. 2013). Although we have attempted to mask out DLA through the DLA catalog provided by the SDSS DR16 data pipeline as in Section 3.1, some of these spectra passed through because the damped Lyα system resides in the Lyα emission. (3) Misclassified quasar spectra by the SDSS DR16 data pipeline. (4) Quasar spectra with erroneous redshift determination. (5) Quasar spectra with damped Lyα absorptions but not flagged by the SDSS DR16 data pipeline. (6) Quasar spectra with an unusually weak Lyα emission. The continuum is almost the same level as the Lyα emission in the example shown. (7) Quasar spectra with BAL around the CIV emission line.

| Model Performance on Wavelengths Longer than the Lyα Emission |
|---------------------------------------------------------------|
| Unperturbed | Perturbed |
| AFFE (%) | 5th | 95th | 5th | 95th |
| QFA versus Truth | 0.56 | 2.11 | 0.35 | 1.71 |
| PCA versus Truth | 0.29 | 4.29 | 1.71 | 6.95 |

Table 7

![Figure 18](image-url) Two examples of quasar continuum inferences on the red side with SDSS DR16 spectra. QFA performs well in both cases. The predicted continua are not affected by the absorption features on the red side and are consistent with the mean flux of the observed SDSS spectra. Conversely, PCA tends to underestimate the red-side continua in both cases due to the inconsistency between the PCA template and the SDSS DR16 spectra.
Appendix F

1D Lyα Forest Power Spectrum Measurement on the Unperturbed Data Set

In Figure 10, we show the 1D Lyα forest power spectrum measurements from the perturbed data set. For completeness, we show the 1D Lyα forest power spectrum measurements in Figure 20. As in the unperturbed data set, three methods give only ~1% difference for continuum prediction, and no significant difference is found for the 1D Lyα forest power spectrum measurements.

Figure 19. Examples of the quasar spectral outliers selected by QFA as described in Section 4.4. The solid gray lines denote quasar spectra, and the orange dashed lines show the flux uncertainties. We visually classified these outliers into seven outlier classes (see Section 4.4 and Appendix E).
Appendix G
Latent Embedding Learned by QFA

In Section 4.5, we focused on three of the most prominent factor components and their dependency on redshift and luminosity. For completeness, Figure 21 shows the other five components learned by QFA. As before (Section 4.5), the components shown are subjected to a varimax rotation.

Unlike PCA analysis (e.g., Figure 2 in Pâris et al. 2011), all eight components (five here and three in the main text) learned by QFA demonstrate clear and prominent features or even singular features. These features include Lyα emission line, power-law slope, and C IV emission line (see Figure 12). Most components are also dominated by visible positive or negative correlation features from two to three features, including Lyα, C IV, N V, O I, C II, Si IV, and C IV emission lines. All components do not show any clear evolution with redshift, demonstrating that the quasar population has not evolved from \( z = 3.5 \) to \( z = 2 \). Also, no evident luminosity dependency is found within these factors except the negative correlation from the particular C IV component, as discussed in the main text.

Figure 20. Similar to Figure 10. We show the 1D Lyα forest measurements on the unperturbed data set (2.9 < \( z < 3.1 \)) here. Three methods give similar uncertainty.
Appendix H

Ablation Study

H.1. Model Dependency on Hidden Dimensions

The number of hidden dimensions, \( N_h \), is a critical hyperparameter in our method. As we described in Section 2, \( N_h \) is much smaller than the dimension of the spectra (\( N_{\text{pix}} \)), and is set to 8 in the main text. Here we perform ablation study to investigate the effect of different \( N_h \) on the model performance. As shown in Table 8, the model remains in good performance when \( N_h \approx 10 \). Too large \( N_h \), e.g., \( N_h = 15 \), will cause the model to overfit some absorption features and degrade the model performance, while too small \( N_h \), e.g., \( N_h = 3 \), may lead to a underfitted model, and also degrade the model performance. From our ablation study, \( N_h \) is appropriate at the range of about 5 ~ 10.

H.2. Model Dependency on Predefined Mean Optical Depth Function

The choice of predefined mean optical depth function is an important prior for our method. As the quasar continuum is intrinsically degenerate with the Ly\( \alpha \) forest, it is unavoidable to assign some prior information to separate the quasar continua from the Ly\( \alpha \) forest reasonably. Although different mean optical depth measurements are generally at the same level, it may be problematic for quasar continuum fitting with a precision requirement of a few percent. Here we investigate the effect of different mean optical depth priors. We will compare the variation of model performance with respect to three literature mean optical depth functions from Faucher-Giguère et al. (2008), Becker et al. (2013), and Kamble et al. (2020).

The mean optical depth functions determined by Faucher-Giguère et al. (2008), Becker et al. (2013), and Kamble et al. (2020) are shown in Figure 22. While the two measurements, Faucher-Giguère et al. (2008) and Becker et al. (2013), from high-resolution and high-S/N quasar spectra show almost consistent result, the measurement (Kamble et al. 2020) from a large number of moderate-resolution and S/N quasar spectra, however, is slightly different. We have adopted the measurement from Becker et al. (2013) in the main text, as described in Section 2.2.

The results in Table 9 indicate that the discrepancies among continuum predictions are commensurate with variations in the mean optical depth function. This suggests that precise measurements of the mean optical depth function are
imperative to disambiguate the quasar continuum and Lyα forest absorption. At present, the mean optical depth function is only employed as an a priori assumption. Therefore, in practice, evaluating alternative mean optical depth functions may be critical.

**H.3. Model Dependency on Regularization Strength**

To impose a prior that our model parameters should be close to zero, we add a regularization term to the log-likelihood function in Section 2.3.2. The strength of the regularization term is controlled by a hyperparameter $\alpha$ in Equation (13). Here, we investigate the dependency of model performance on $\alpha$. As shown in Table 10, too small $\alpha$ will not provide enough constraints and will lead to poor performance, while too large $\alpha$ will make the variations of model parameters too small, thereby weakening the model’s performance. $\alpha$ is appropriate at the level of $10^{-1}$.

![Figure 22. Mean optical depth measurements from Faucher-Giguère et al. (2008), Becker et al. (2013), and Kamble et al. (2020). We have adopted the result from Becker et al. (2013; see Equation (3)) in the main text.](image)

| $N_h$ | 5 | 6 | 7 | 8 | 9 | 10 | 15 |
|-------|---|---|---|---|---|----|----|
| AFFE [%] (50th) | 2.43 | 2.44 | 2.47 | 2.47 | 2.48 | 2.65 | 4.41 |

**Table 9**

| $\tau_{\text{eff}}$ | Faucher-Giguère et al. (2008) | Becker et al. (2013) | Kamble et al. (2020) |
|---------------------|-------------------------------|---------------------|---------------------|
| AFFE [%] (50th)     | 2.15                          | 2.48                | 5.23                |

**Note.** Variations in continuum modeling resulting from different choices of mean optical depth are on par with the differences between the mean optical depths themselves. In application, adopting alternative mean optical depth measurements may prove necessary.

![Table 8](image)

![Table 10](image)
Appendix I
Underestimation of Continuum Uncertainty and Possible Calibrations

As discussed in Sections 2.4 and 5.2, omitting the stochastic error term $\Psi$ during inference is necessary to optimize the model’s performance in estimating the posterior mean of the continua, as certain erratic features tend to be subsumed into this term. However, such an approach will significantly underestimate the uncertainty in QFA. As illustrated in Figure 23, by performing an analogous experiment as Reiman et al. (2020), which assesses the performance of the observed confidence intervals by calculating the fraction of observed absolute errors that fall within the $1\sigma$ (68%) and $2\sigma$ (95%) confidence intervals under the Gaussian assumption, we concluded that the empirical error observed is appreciably greater than the inferred uncertainty. However, if we scale the inferred uncertainty by a factor of 3, we find that the inferred uncertainty closely matches the empirical uncertainty, as demonstrated in Figure 24.

Figure 23. Uncalibrated observed confidence intervals for QFA as a function of rest-frame wavelength (similar to Figure 11 in Reiman et al. 2020). If the uncertainty given by QFA were calibrated accurately, $P\%$ of the observed absolute errors would fall within the $P\%$ confidence interval. Under the Gaussian assumption of QFA, we show here the observed confidence intervals for $1\sigma$ and $2\sigma$ (e.g., $P = 68\%$ and $P = 95\%$). The solid horizontal lines correspond to the expected confidence intervals. QFA produces overly optimistic confidence intervals because of the omission of the stochastic error term $\Psi$.

Figure 24. Similar to Figure 23, but scaling the inferred uncertainty to a factor of 3. In that case, the observed and inferred uncertainty values exhibited close correspondence.
