The Astrophysical Journal, 938:17 (13pp), 2022 October 10 https://doi.org/10.3847/1538-4357/ac8ead
© 2022. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

A Generative Model for Quasar Spectra

Anna–Christina Eilers[1,7], David W. Hogg[2,3,4], Bernhard Schölkopf[5], Daniel Foreman-Mackey[4], Frederick B. Davies[3], and Jan–Torge Schindler[3,6]
[1] MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; eilers@mit.edu
[2] Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA
[3] Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany
[4] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA
[5] Max Planck Institute for Intelligent Systems, Tübingen, Germany
[6] Leiden Observatory, Leiden University, Niels Bohrweg 2, NL-2333 CA Leiden, The Netherlands
Received 2022 July 6; revised 2022 August 31; accepted 2022 August 31; published 2022 October 11

Abstract

We build a multi-output generative model for quasar spectra and the properties of their black hole engines, based on a Gaussian process latent-variable model. This model treats every quasar as a vector of latent properties such that the spectrum and all physical properties of the quasar are associated with non-linear functions of those latent parameters; the Gaussian process kernel functions define priors on the function space. Our generative model is trained with a justifiable likelihood function that allows us to treat heteroscedastic noise and missing data correctly, which is crucial for all astrophysical applications. It can simultaneously predict unobserved spectral regions and the physical properties of quasars in held-out test data. We apply the model to rest-frame ultraviolet and optical quasar spectra for which precise black hole masses (based on reverberation-mapping measurements) are available. Unlike reverberation-mapping studies that require multi-epoch data, our model predicts black hole masses from single-epoch spectra—even with limited spectral coverage. We demonstrate the capabilities of the model by predicting black hole masses and unobserved spectral regions. We find that we predict black hole masses at close to the best possible accuracy.

Unified Astronomy Thesaurus concepts: Gaussian Processes regression (1930); Nonparametric inference (1903); Quasars (1319); Astrostatistics techniques (1886); Spectroscopy (1558); Supermassive black holes (1663)

1. Introduction

Machine-learning algorithms mainly fall into the two categories of *supervised* and *unsupervised*. In supervised learning tasks, data points that consist of features and labels are used to train the model, such that it can classify data into different categories (when the labels are discrete), or model a continuous relationship between features and labels (when the labels are real-valued or lists of real values). Unsupervised learning algorithms generally aim to understand the structure of a data set, uncover patterns in the data, or cluster unlabeled data sets. An equally important distinction in machine learning is made between *discriminative* models and *generative* models. Discriminative models are designed to find functions of data features that predict data labels. Generative models are designed to find functions that predict features, or predict features given labels, or predict both features and labels.

In the natural sciences—and astrophysics in particular—generative models have an advantage over discriminative models because they can naturally include any peculiar uncertainties, noise, and missing-data properties in the data set. This is possible because the generative model synthesizes the features; it can be trained with a justifiable loss function such as a log-likelihood that contains a reasonable representation of the noise model in the feature space, and drops missing data. This is in contrast to discriminative models, most of which require complete, rectangular data, which are implicitly believed to be correct or true.

One approach to generative, unsupervised machine learning is to represent a complex high-dimensional data set in a lower-dimensional latent space. The archetype of this method is principal component analysis (PCA), which seeks a linear projection of the data onto a lower-dimensional subspace, represented by an orthonormal basis, such that the data generated from the low-dimensional PCA preserves as much variance as possible from the original data. Because the input data can be reconstructed from the PCA projections, there is a sense in which PCA can be thought of as a generative model. Since PCA is linear, it is often incapable of representing the full structure in the data through its linear low-dimensional embedding.

A Gaussian process latent-variable model (GPLVM) is a fully probabilistic, non-linear, and generative latent-variable model that generalizes PCA (Lawrence 2003, 2005; Lawrence & Moore 2007; Titsias & Lawrence 2010). It is a generative model that represents a non-linear extension of the linear probabilistic PCA (Tipping & Bishop 1999) and has been shown to be a powerful approach for probabilistic modeling of high-dimensional data through dimensionality reduction. The classical GPLVM is unsupervised—it does not distinguish features and labels and is not designed for supervised learning tasks.

In this work, we aim to introduce and provide a framework to apply the GPLVM to astrophysical settings. In particular, we are interested in problems that come with noisy, heteroscedastic, and complex data sets that contain both spectral features and labels, where some parts of the data set might be missing or

[7] NASA Hubble Fellow.

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

unobserved. To this end, we construct a modified version of the standard GPLVM and expand the algorithm to a "multi-output" generative model (see also Gao et al. 2011). This multi-output GPLVM simultaneously generates both the features and their associated labels from a common latent space (see Section 2), and thus enables predictions of both the spectral features and the labels.

As a first application (see Section 5), we apply this generative model to quasar spectra aiming to determine their physical properties, such as the masses of their central accreting supermassive black holes (SMBHs), based on their single-epoch spectra alone. Determining accurate black hole masses for quasars is challenging and usually requires time-intensive, multi-epoch observations with regular cadence to conduct reverberation-mapping (RM) measurements (e.g., Peterson 1993; Barth et al. 2015; Shen et al. 2016), which are currently unfeasible for quasars beyond redshift $z \gtrsim 2$. Thus, for the vast majority of quasars, black hole mass estimates are obtained by means of scaling relations that relate the quasar's luminosity and the emission line widths observed in their spectra to black hole mass estimates that are calibrated based on low-redshift reverberation-mapped quasars (e.g., Vestergaard & Peterson 2006; Coatman et al. 2017; Grier et al. 2017). However, we know that information about the quasars' black hole masses is encoded in the single-epoch quasar spectra, which we aim to reveal by means of our generative model.

In the chosen generative model that we present here, both the quasar spectra and their physical labels (such as their black hole masses) are simultaneously generated from points in a latent space. Because the data set contains missing (or unobserved) data (and because the non-missing data have heteroscedastic noise properties), the generative model has the advantage that every extant piece of data can be handled appropriately, and missing features and labels are no problem. The generative model can also correctly account for different data measurement precisions (i.e., data weights).

The single biggest issue with the model is the small size of the available training data. There are currently only 31 quasars that meet our data-quality cuts. Despite this, th.e model makes good predictions for black hole masses and spectral pixels in held-out data. We will discuss this and other limitations of the model in Section 6, and highlight possible future applications and improvements in Section 7.

2. A Gaussian Process Latent Variable Model

The GPLVM is a generative model that represents a flexible, non-linear approach for a dimensionality reduction using a Gaussian process to learn a low-dimensional representation of (potentially) high-dimensional data. Typically, the GPLVM is used for completely non-supervised learning tasks but here we modify the algorithm to a multi-output generative model, such that it can simultaneously generate *both* the data features and the associated labels. Furthermore, the GPLVM can naturally handle data sets with heteroscedastic uncertainties and missing or unobserved data, which is crucial for any application to real astronomical problems. However, dealing with heteroscedastic noise and missing data implies that the implementation of the GPLVM becomes significantly more complicated and less scalable to large data sets, which we discuss further in Section 6.

2.1. Assumptions

Our model makes a number of strong assumptions. In particular, we assume that our data is correct, in the sense that the measurements are unbiased, with normally distributed noise, and do not include substantial outliers. Relatedly, we assume that all measurements (i.e., spectral pixels and labels) are independent measurements with no (or negligible) covariances (although we will later weaken this assumption in Equation (7)). We assume that the high-dimensional data set in the joint space of labels and features (spectral pixels) is drawn from a distribution that is intrinsically low dimensional in nature, such that each data point can be represented as a point in a lower-dimensional latent space. Additionally—and importantly—we assume that the mapping of the latent space to observations can be expressed as draws from a Gaussian process.

2.2. Input Data

Let us assume that we have N training-set objects. Their features (e.g., quasar spectral pixels) make up a $N \times D$ rectangular matrix X (this matrix will however, have many missing values, see Section 2.3), which are given as

$$X = [\vec{x}_1, ..., \vec{x}_N]^T \text{ with uncertainties } \sigma_X = [\vec{\sigma}_{x_1}, ..., \vec{\sigma}_{x_N}]^T \ , \quad (1)$$

where the $\vec{x}_n$ are individual D-vector spectra. Associated with these features is an $N \times L$ rectangular matrix Y of labels

$$Y = [\vec{y}_1, ..., \vec{y}_N]^T \text{ with uncertainties } \sigma_Y = [\vec{\sigma}_{y_1}, ..., \vec{\sigma}_{y_N}]^T \ , \quad (2)$$

where the $\vec{y}_n$ are individual L-vector labels. In our example application in Section 5, the labels include the black hole mass, bolometric luminosity, and redshift of the quasars, but other physical properties can easily be added as additional labels. Technically, the uncertainty information could be full covariance matrices but in this case we will treat the uncertainties as independent, such that the uncertainties can be represented with objects the same sizes as X and Y.

For a stable optimization of the model, we re-scale the input data set, such that each of the D columns of X and the L columns of Y have zero mean and unit variance. The $\vec{x}_n$ and $\vec{\sigma}_{x_n}$ are scaled consistently, as are the $\vec{y}_n$ and the $\vec{\sigma}_{y_n}$.

2.3. Handling Missing Data

Although the input data X and Y are technically rectangular (i.e., two-dimensional matrices of shape $N \times D$ and $N \times L$, respectively, where every record has the same length and the same feature structure), in practice there is a lot of missing data. For instance, some labels might not be measured for a subset of objects, or the spectra in the data set might be observed with different telescopes or instruments, resulting in a different wavelength coverage. Additionally, we will transform all quasars to their rest-frame wavelengths (see Section 5.2). Thus, even quasars observed with the same telescope and instrument but at slightly different redshifts will also have a different rest-frame wavelength coverage. These missing data entries yield to sparse data matrices X and Y.

The standard GPLVM can handle the missing data in a conceptually rigorous way by only taking objects into account that have finite data (i.e., measured values) at any given pixel d or label l. In this way, objects with missing or unknown labels l can also be accounted for in the training set. The missing data

and uncertainties in the rectangular input data are represented with NaNs.

2.4. Kernel Functions

The idea of the GPLVM is that the state of object n can be represented with a Q-dimensional latent vector $\vec{z}_n$. These can be combined into a rectangular $N \times Q$ latent-variable block Z

$$Z = [\vec{z}_1, ..., \vec{z}_N]^T. \tag{3}$$

This one set of latents will generate all of the spectra and all of the labels, or in other words the Q-vector $\vec{z}_n$ will generate the spectrum $\vec{x}_n$ and labels $\vec{y}_n$ of object n.

The Gaussian processes are defined by kernel functions; the kernel functions determine the prior over functions. For the kernel function relating the features X to the latents, we choose the commonly used radial basis function (RBF) kernel

$$K_x(z_i, z_j) = A_x \exp\left[-\frac{B_x}{2}(z_i - z_j)^T(z_i - z_j)\right], \tag{4}$$

while we choose different hyperparameters but the same kernel function relating the labels Y to the latents; that is,

$$K_y(z_i, z_j) = A_{y,l} \exp\left[-\frac{B_y}{2}(z_i - z_j)^T(z_i - z_j)\right]. \tag{5}$$

The hyperparameters A_x and $A_{y,l}$ constitute the amplitude of the kernels, while the hyperparameters B_x and B_y denote the length scales of the kernels. Note that there is a redundancy between the length scales and the overall scale of the latent variables Z, which is ameliorated by adding a prior on the latent parameters (third term in Equation (9)). However, in practice we find that we obtain a better optimization of the model when keeping the scale lengths B_x and B_y simply fixed to unity. Note that we take only one kernel function for all D-dimensional features (Equation (4)), while we use a different kernel for each label l (Equation (5)) (i.e., the hyperparameter $A_{y,l}$ can be different for each label l). In what follows, we will refer to the set of hyperparameters as $\theta = [A_x, A_{y,l=0}, ..., A_{y,l=L}]$.

2.5. Accounting for Heteroscedasticity in the Input Data

The quasar spectra X and the labels Y have measurement uncertainties σ_X and σ_Y, respectively, which can naturally be accounted for in a GPLVM. To account for the heteroscedasticity in the data, we construct covariance matrices for each spectral pixel d and each label l; that is,

$$C_d = \begin{bmatrix} \sigma^2_{x_{1,d}} & 0 & ... & ... \\ 0 & \sigma^2_{x_{2,d}} & 0 & ... \\ \vdots & \vdots & \ddots & \vdots \\ ... & ... & 0 & \sigma^2_{x_{N,d}} \end{bmatrix} \text{ and } C_l = \begin{bmatrix} \sigma^2_{y_{1,l}} & 0 & ... & ... \\ 0 & \sigma^2_{y_{2,l}} & 0 & ... \\ \vdots & \vdots & \ddots & \vdots \\ ... & ... & 0 & \sigma^2_{y_{N,l}} \end{bmatrix} \tag{6}$$

that will be added to the kernel functions, i.e.,

$$\hat{K}_{x,d} = K_x + (1 + \beta)C_d \tag{7}$$

$$\hat{K}_{y,l} = K_y + C_l. \tag{8}$$

The term $(1 + \beta)$ has been added to capture the "information content" of the spectra. If we completely trusted the noise model in our quasar spectra, then we would set $\beta = 0$, while a higher value of β indicates that the noise in the quasar spectra might be underestimated. Thus, in practice β represents another hyperparameter to our model. Note that this represents a tractable solution for capturing the covariant structure of the spectra, but a much better and more correct approach would be to use a Gaussian process in the spectral dimension. However, given the structure of our model, this would be extremely difficult to make computationally tractable and is thus beyond the scope of this paper.

The kernel matrices $\hat{K}_{x,d}$ and $\hat{K}_{y,l}$ have dimensions $N_d \times N_d$ and $N_l \times N_l$, respectively, where N_d and N_l denote the number of objects in the training set that have a measured and finite data point at pixel d or label l.

2.6. Likelihood Functions

The logarithm of the joint probability of the features X, the labels Y, and the latent variables Z in the GPLVM represents the "cost function" and is given as

$$\ln \mathscr{L}(X, Y|\theta, Z) = \ln \mathscr{L}_x(X|\theta, Z) + \ln \mathscr{L}_y(Y|\theta, Z) + \ln \mathscr{P}_z(Z) \tag{9}$$

The first term in Equation (9) denotes the likelihood function for the input features X, which includes a sum over all pixels, assuming that all pixels can be treated independently. The likelihood function is conditioned on the latent variables Z and the hyperparameters θ , i.e.

$$\ln \mathscr{L}_x(X|\theta, Z) = \sum_{d=1}^{D}\left[-\frac{N_d}{2}\ln 2\pi - \frac{1}{2}\ln(\det \hat{K}_{x,d}) - \frac{1}{2}(X_{:,d}^T\, \hat{K}_{x,d}^{-1}\, X_{:,d})\right], \tag{10}$$

where $X_{:,d}$ denotes the input features of all quasars at pixel d, and N_d denotes the number of objects with given finite input data at pixel d.

Analogously, the second term in Equation (9) describes the likelihood function for the labels Y, conditioned on the latent variables Z and the hyperparameters θ; that is,

$$\ln \mathscr{L}_y(Y|\theta, Z) = \sum_{l=1}^{L}\left[-\frac{N_l}{2}\ln 2\pi - \frac{1}{2}\ln(\det \hat{K}_{y,l}) - \frac{1}{2}(Y_{:,l}^T\, \hat{K}_{y,l}^{-1}\, Y_{:,l})\right]. \tag{11}$$

Here, $Y_{:,l}$ indicate the label values of label l for all objects, while N_l denotes the number of objects with given label l.

The third term in Equation (9) constitutes a *prior* on the latent variables Z, for which we choose a Gaussian with zero mean and unit variance, i.e.,

$$\ln \mathscr{P}_z = -\frac{N}{2}\ln 2\pi - \frac{1}{2}Z^2. \tag{12}$$

As mentioned previously, this prior is useful in alleviating the redundancy between the latent variables and the scale lengths B_x and B_y. Nevertheless, there are still many exact degeneracies between the different latent dimensions, in the sense that this model is rotationally symmetric in the latent space Z.

Note that Equations (10) and (11) follow the same mathematical structure and thus could be combined to one likelihood function treating the labels as a second set of features. Similarly, one could split Equation (11) into multiple separate likelihood functions for each individual label l, while mathematically the structure of our model would still remain unchanged. We consider this GPLVM, where the latent space generates multiple different outputs (in this case, the spectral features of quasars as well as their physical labels), to be a "multi-output" generative model.

In contrast to standard Gaussian process regression where one fits for the data, our model optimizes for the latent parameters Z and also the hyperparameters A_x and $A_{y,l}$ that maximize the cost function in Equation (9) for the data. Because the X- and Y-space predictions are non-linear functions of the latents Z, the optimizations are non-linear and are not guaranteed to find a global optimum. Note that technically this model is not fully Bayesian because at training time parameters are found by optimization.

3. Model Predictions

3.1. "z-step": Finding a Set of Latent Parameters for New Input Data

Once we have optimized the GPLVM with respect to the latent parameters Z and also the set of hyperparameters θ, we can apply it to a new and yet unseen data point with spectrum x_*,[8] which could have a subset of already known labels y_*, to predict its set of unknown labels $\tilde{y}_*$ and possibly missing or unobserved spectral pixels $\tilde{x}_*$. We first find (by optimization) the set of latent parameters z_* that represents the new input $\{x_*, y_*\}$, which will then be used to estimate the unknown part of the data.

A benefit of Gaussian processes (e.g., Williams 1998; Lawrence 2005; Rasmussen & Williams 2005) is that any new data point $\{x_*, y_*\}$ (with measurement uncertainties $\{\sigma_{x_*}, \sigma_{y_*}\}$) that is not part of the training set will have a posterior probability distribution function (PDF) that is a Gaussian; that is,

$$\mathscr{L}(x_*, y_*|z_*) = \mathcal{N}(x_*, y_*|\mu_x(z_*), s_x^2(z_*), \mu_y(z_*), s_y^2(z_*)), \tag{13}$$

with means μ_x and μ_y, and variances s_x^2 and s_y^2 (given in Equations (15) to (18) below).

The set of latent variables z_* for the new data point $\{x_*, y_*\}$ is determined by computing the likelihood of the observed data given the projection of the posterior probability estimate for z_* back into the data-space, i.e.,

$$\begin{aligned}\ln \mathscr{L}(x_*, y_*|z_*, X, Y, Z, \theta) = &-\frac{1}{2}\ln(s_x^2 + \sigma_{x_*}^2) \\ &- \frac{1}{2}\frac{(x_* - \mu_x)^2}{s_x^2 + \sigma_{x_*}^2} - \frac{N_d}{2}\ln(2\pi) \\ &- \frac{1}{2}\ln(s_y^2 + \sigma_{y_*}^2) - \frac{1}{2}\frac{(y_* - \mu_y)^2}{s_y^2 + \sigma_{y_*}^2} - \frac{N_l}{2}\ln(2\pi)\end{aligned} \tag{14}$$

with a mean for the feature space X

$$\mu_x = X^T \hat{K}_x^{-1} k_x(Z, z_*) \tag{15}$$

and variance

$$s_x^2 = k_x(z_*, z_*) - k_x(Z, z_*)^T \hat{K}_x^{-1} k_x(Z, z_*), \tag{16}$$

and analogously for the label space Y; that is,

$$\mu_y = Y^T \hat{K}_y^{-1} k_y(Z, z_*) \tag{17}$$

and

$$s_y^2 = k_y(z_*, z_*) - k_y(Z, z_*)^T \hat{K}_y^{-1} k_y(Z, z_*). \tag{18}$$

Here, $k_x(Z, z_*)$ and $k_y(Z, z_*)$ denote the column vector constructed from computing the elements of the kernel matrices between the training set and the new point in latent space z_* (see e.g., Williams 1998; Lawrence 2005; Rasmussen & Williams 2005, for details).

Note that the posterior distribution over Z could be multi-modal with respect to z_*. Thus one should use sampling methods to evaluate the posterior distribution and then approximate z_* around the largest mode. However, in practice, we do not find any multi-modality for the chosen application (see Section 5), and hence simply take the set of parameters z_* that maximizes Equation (14).

3.2. Predicting Unknown Labels $\tilde{y}_*$ and Missing Spectral Pixels $\tilde{x}_*$

Once we have a set of latent variables z_* that represent the new data point $\{x_*, y_*\}$, the prediction for the *unknown* labels $\tilde{y}_*$ also follows a Gaussian distribution (e.g., Lawrence 2005); that is,

$$\mathscr{L}(\tilde{y}_*|z_*, Y, Z, \theta) = \mathcal{N}(\tilde{y}_*|\mu_y(z_*, Y, Z, \theta), s_y^2(z_*, Z, \theta)) \tag{19}$$

with a mean and variance given in Equations (17) and (18). Note that this step does not require any further optimization.

Given the set of latent variables z_* that represents the new quasar in the latent space, we can not only predict its unknown labels but can also predict its missing spectral pixels $\tilde{x}_*$ because the latent variables in our model represent both the features (i.e., spectra) and also the labels. Note that technically this is exactly the same as predicting the unknown labels $\tilde{y}_*$ in Equation (19), and hence each (missing) spectral pixel of an object, which can either be part of the training set or a new unseen spectrum, is given by

$$\mathscr{L}(\tilde{x}_*|z_*, X, Z, \theta) = \mathcal{N}(\tilde{x}_*|\mu_x(z_*, X, Z, \theta), s_x^2(z_*, X, Z, \theta)), \tag{20}$$

where μ_x and s_x^2 are given by Equations (15) and (16).

[8] Note that we omit the vector notation here for better readability, but please keep in mind that throughout this manuscript the testing objects naturally have the same dimensions as the training-set data (i.e., x_* is a D-dimensional vector and y_* is an L-dimensional vector).

3.3. Predicting Spectral Features for Given Labels: Sampling the Latent Space

Just as it is possible to predict missing pixels from the feature vector, it is also possible to predict the entire feature vector given an input label. However, the problem with this is that the assumption that the data are uniquely represented as a point in the latent Z-space becomes untrue if the only input to the test-step likelihood is a single label and no spectrum. Thus, to understand how quasar spectra depend on their labels, such as the black hole mass, we sample the Q-dimensional latent space and find regions that correspond to sets of latent parameters representing quasars with the given input label.

In practice, we take Gaussian random draws from the Q-dimensional latent space and determine the corresponding label y_* using Equation (19). We then find regions in the latent Z space that represent the same label y_* and determine their spectral features using Equation (20). By averaging the spectral features from many latent representations of the same label y_*, we can search for spectral dependencies on the labels. As a first example, we will show later how quasar spectra depend on their physical properties (see Figure 6 in Section 5).

4. Implementation Notes

We use the L-BFGS-B algorithm (Zhu et al. 1997) to optimize the cost function in Equation (9) of our model. We simultaneously optimize the Q-dimensional latent parameters Z for each object in our training set, as well as for the hyperparameters describing the amplitudes of the kernel functions A_x and $A_{y,l}$, where the latter can be different for each label l. Thus, in practice we are optimizing $N \cdot Q + (L+1)$ parameters. We take the derivatives analytically and include the Jacobian matrix for better optimization.

The latent dimension Q and the hyperparameter β from Equation (7) are optimized via cross-validation. Since the main goal of our first application described in Section 5 will be to predict the black hole masses from single-epoch quasar spectra, we train the GPLVM N times on $N-1$ quasars and afterwards predict the black hole mass label of the omitted Nth object. This procedure is then repeated for different values for Q and β. We chose a set of parameters Q and β that minimizes the bias and scatter in the distribution of predicted black hole masses compared to the input black hole masses in this cross-validation (see Figure 3 in Section 5).

Our model permits (and requires!) many choices regarding not only the optimization criteria for hyperparameters but also the input and output parameters. Furthermore, the choice of the kernel functions (Equations (4) and (5)) is a strong model assumption and can be optimized by cross-validation. The choices of the model parameters and kernel functions will depend on the specific goals and applications of the GPLVM. When we apply our model to quasar spectra in the next section, we will make different choices with regards to the number L and chosen labels when predicting the quasars' black hole masses (see Figure 3) or when understanding the spectral dependencies of quasar spectra on various physical properties (see Figure 6). Note that we do not conduct an extensive parameter search to find the "best" model choices for our application, and thus the model might perform better with a different and more optimized set of parameters for a given application and science goal.

We will discuss the limitations of our current model implementation and some advanced algorithms that will significantly improve the performance of our GPLVM in the future in detail in Section 6.

5. A First Application: Predicting the Physical Properties of Quasars from Their Spectra

Our multi-output generative GPLVM has numerous potential applications. Here, we want to provide a first example, where we apply the model to single-epoch quasar spectra to predict black hole mass measurements from their spectral features alone. We briefly summarize how SMBHs are measured in quasars to give some context (Section 5.1), before introducing our data set of quasar spectra for which the masses of their central SMBHs are well known (Section 5.2). We will show that the prediction accuracy of the SMBH masses from the GPLVM is as good as the measurements allow (Section 5.3) and further show how our model can predict missing or unobserved spectral pixels (Section 5.4). At the end, we demonstrate how the quasar spectra depend on their physical parameters (Section 5.5).

5.1. Context: Measuring the Masses of Supermassive Black Holes

The masses of the central SMBHs of nearby galaxies can be measured by resolving the sphere of influence around the black hole in the motion of stars or gas, which has resulted in the now well-established $M_\bullet - \sigma_*$ relation at $z \sim 0$ (e.g., Magorrian et al. 1998; Gebhardt et al. 2000; Häring & Rix 2004; Gültekin et al. 2009) relating the mass of a black hole $M_\bullet$ in the center of galaxies to the stellar velocity dispersion σ_* in the galactic bulge. However, this is not feasible for quasars because the central accreting black hole outshines the stellar light by several orders of magnitudes.

For a few quasars at low redshifts (i.e., $z \lesssim 1$), precise mass estimates have been obtained via the so-called RM technique or *echo mapping* (e.g., Blandford & McKee 1982; Peterson 1993). This enables measurements of the distance between the SMBH in the center of a quasar and gas clouds orbiting the black hole within the so-called broad-line region (BLR). Assuming that the gas motion in the BLR is completely dominated by the gravitational pull of the black hole, one can then derive the black hole mass using Newton's law of motion; that is,

$$M_\bullet = f \frac{R_{\rm BLR}\ \Delta v^2}{G} = f\ \frac{c\ \tau\ \sigma_{\rm rms}^2}{G}, \tag{21}$$

where Δv denotes the width of the emission lines broadened by Doppler broadening due to the velocity of the orbiting gas clouds (which can be inferred from the variance $\sigma_{\rm rms}^2$ of the emission lines in the spectra of the quasars), G is the gravitational constant, and f is a geometric factor to account for the unknown geometry and gas distribution of the BLR. The radius of the BLR, $R_{\rm BLR}$, is estimated by means of the RM method, which measures the time lag τ between changes in the continuum emission arising from the accretion disk around the black hole and the corresponding line emission changes from the gas clouds, once the radiation has propagated outwards from the black hole to the BLR. Thus, the radius of the broad-line region can be estimated as $R_{\rm BLR} = c\tau$, with c as the speed of light.

The average geometric factor f is determined for an ensemble of objects by enforcing that the black hole mass measurements fall on the well-known local $M_\bullet - \sigma_*$ relation (e.g., Onken et al. 2004). However, this has an intrinsic scatter of 0.35–0.5 dex (e.g., Magorrian et al. 1998; McLure & Dunlop 2002), and therefore represents the limiting precision for all black hole mass measurements. Additional uncertainties can arise if there is a redshift evolution of the $M_\bullet - \sigma$ relation (e.g., Pensabene et al. 2020) or if quasars were to obey a different scaling relation than quiescent galaxies without nuclear activity (e.g., Woo et al. 2015).

Since the RM measurements require long monitoring and expensive observations, this method is unfeasible for most quasars—especially at higher redshifts where time delays are longer due to time dilation and the generally more massive black holes. Thus, masses of SMBHs for most quasars are commonly inferred from single-epoch spectra by using scaling relations that relate the width of an emission line and the quasar's luminosity to the black hole mass (e.g., Vestergaard & Peterson 2006; Grier et al. 2017; Coatman et al. 2017). These scaling relations are calibrated based on quasars with precise RM black hole mass measurements. However, additional uncertainties arise for quasars at high redshifts beyond $z \gtrsim 3$ for two reasons: first, the high-redshift quasars have generally more massive black holes and are more luminous than the observed quasar sample at lower redshift, which requires an extrapolation of the scaling relations in a parameter space that is only sparsely sampled at lower redshifts; and second, because the rest-frame optical emission lines (e.g., H β) that are commonly used for calibrating the scaling relations can no longer be measured with ground-based observatories at high redshifts, additional scaling relations between the width of the Hβ emission line and rest-frame UV lines (e.g., Mg II or C IV,) are required to estimate the black hole masses (e.g., Wang et al. 2009). These various scaling relations and extrapolations make black hole mass estimates for most quasars in the universe highly uncertain and potentially biased.

With the generative model presented here, we aim to circumvent all scaling relations and intend to directly constrain the black hole masses of quasars (and other physical properties) from the single-epoch spectra themselves.

5.2. Training Data: Quasar Spectra with Precise Black Hole Mass Measurements

Our training data set consists of 31 quasar spectra for which reliable RM black hole mass measurements have been reported in the literature (e.g., Bentz et al. 2009; Barth et al. 2015). We chose only quasars with reliable Hβ emission line time lags due to the large scatter in the measured time lags observed between different emission lines (e.g., Fausnaugh et al. 2017; Grier et al. 2017).

The quasar spectra are obtained with two different instruments—the Space Telescope Imaging Spectrograph (STIS) and the Cosmic Origins Spectrograph (COS) on the Hubble Space Telescope (HST)—, covering the rest-frame UV and optical wavelengths (Park et al. 2013, 2017). We transform all spectra to rest-frame wavelengths and take all observed spectral pixels between 1220–5000 Å into account. This wavelength range is chosen to avoid any absorption from the intergalactic medium (IGM) bluewards of the Lyα emission, and to include the Hβ emission line at $\lambda_{\rm rest} \approx 4861$ Å. Due to the different redshifts of the quasars (i.e., $0.002 \leqslant z \leqslant 0.234$), all quasars cover slightly different rest-frame wavelengths. Thus, the data set is highly heteroscedastic with a lot of missing data and also varying data quality (i.e., with signal-to-noise ratios between $5 \lesssim {\rm S/N} \lesssim 107$).

To prepare the rectangular input data, we apply a cubic spline fit to each quasar spectrum and afterwards apply a 3σ-clipping to mask any absorption lines within the quasar continua, which arise due to intervening foreground absorption systems along the quasar sightline and are thus not intrinsic to the quasars themselves. We then fit a power-law continuum—that is, $f_\lambda \propto \lambda^{-\alpha}$—to the spectral regions free of emission lines in the quasar spectra and re-scale all spectra to be approximately unity at $\lambda_{\rm rest} = 2500$ Å by dividing the spectra by the value of the power-law continuum at this wavelength. We bin all spectra to a common wavelength grid between $1220\,\mathring{A} \leqslant \lambda \leqslant 5000\,\mathring{A}$ with a fixed pixel scale of $\Delta\lambda = 2$ Å without correlating the noise of neighboring pixels. All pixels that are unobserved or have been masked are set to NaNs. Furthermore, for a more stable optimization, we pivot and scale all pixels to have a mean of zero and unity variance. This results in a rectangular input data set X of shape $N \times D$. The matrix σ_X contains the corresponding measurement uncertainties on each pixel value.

In this first example, we choose either two or three labels for each quasar: we always take the quasars' black hole masses $M_\bullet$ and bolometric luminosities $L_{\rm bol}$ as labels, and for some applications we also take the quasars' redshifts z into account (e.g., Figure 6). However, the label vector could easily be augmented by additional quasar properties, such as, for example, the Eddington ratio of their mass accretion rate $\lambda_{\rm Edd}$. All labels are measured for our chosen data set, but in principle any missing or unknown labels could be set to NaNs. Thus, we could include objects in the training set for which we do not have (or do not wish to include) certain label measurements. Finally, we construct two rectangular matrices Y and σ_Y of shape $N \times L$ containing the input labels and their uncertainties, respectively.

The uncertainties on the redshift z arise solely from measurement uncertainties of the peak of the Hβ emission line that is used to derive the quasars' redshifts. Uncertainties on $L_{\rm bol}$ also only contain measurement uncertainties of the monochromatic luminosity λL_{1350}, which we transform into a bolometric luminosity using the bolometric correction factor of 4.3 (Richards et al. 2006; Vestergaard & Osmer 2009). The uncertainty in the black hole mass estimates arises from a combination of factors: first, we have measurement uncertainties in the time lag τ, as well as the line width $\sigma_{\rm rms}$; and second, the dominating uncertainty in the black hole mass measurements arises from the geometric (or virial) factor f, which relates the measured virial product to the black hole mass estimates by calibrating the measurements to the local $M_\bullet - \sigma_*$ relation (e.g., Onken et al. 2004; Woo et al. 2015). We use a recent measurement by Woo et al. (2015) for the virial factor—that is, $\log_{10} f = 0.65 \pm 0.12$—, which the authors derive by jointly fitting the $M_\bullet - \sigma_*$ relation using local quiescent galaxies and reverberation-mapped active galactic nuclei. Using this virial factor, we update the black hole mass measurements for our data sample that were reported in Park et al. (2013, 2017).

Note that the virial factor f can only be determined for an ensemble of galaxies and quasars by requiring that the black hole mass estimates fall *on average* onto the local $M_\bullet - \sigma_*$

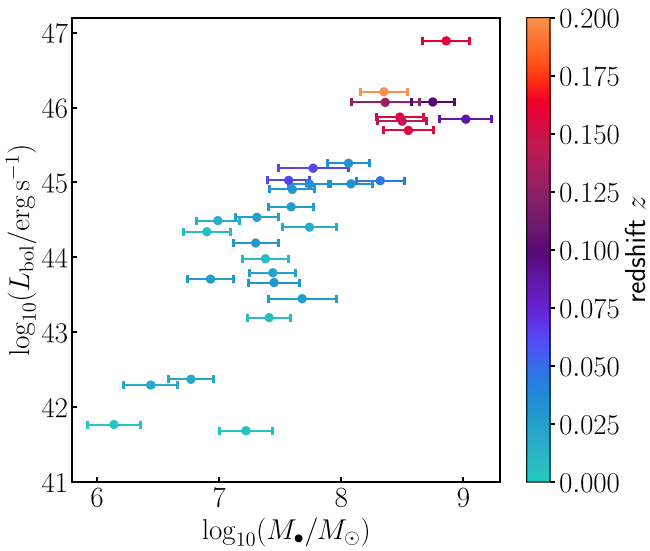

Figure 1. Back hole masses and bolometric luminosities colored by redshift of 31 quasars in our data sample, which have reliable black hole mass measurements based on the RM technique.

relation. Thus, all black hole mass measurements—even when precisely determined via RM measurements—have an intrinsic scatter of approximately 0.35–0.5 dex (e.g., Vestergaard & Peterson 2006; Vestergaard & Osmer 2009; Woo et al. 2015; Park et al. 2017), which constitutes a systematic uncertainty on all black hole mass measurements. This intrinsic scatter limits the precision of the black hole mass estimates that we can possibly achieve.

The properties of the quasars in our data set are shown in Figure 1 and the spectra are shown in Figure 2.[9]

5.3. Predicting Black Hole Masses from Single-epoch Quasar Spectra

As described in Section 4, we train the GPLVM with different parameters for the latent dimension Q and the hyperparameter β, and determine the best values for these parameters by cross-validation of the predicted black hole masses compared to the input black hole mass labels. We find the best predictions for the black hole masses for values of $Q = 16$ and $\beta = 10$, as shown in Figure 3. The predicted black hole mass labels are nearly unbiased with an offset of $\sim$1% and have a scatter of approximately 0.4 dex, which is en par with the best possible precision that we can achieve because it agrees with the intrinsic scatter of the RM measured black hole masses (e.g., Vestergaard & Peterson 2006; Woo et al. 2015).

While the uncertainties on the measured input black hole masses are all comparable in size, because the uncertainty on the virial factor f dominates the error budget, the errorbars on the predicted black hole masses reflect the "information content" of the input data (i.e., the sizes of the errorbars correlate somewhat with the signal-to-noise ratio of the spectrum) and also with the spectral coverage of the input spectrum.

[9] Note that we excluded one object (3C390) from the data set presented in Park et al. (2017) due to its very unusual emission line shape, which was caused either by a strong N IV] λ1486Å emission line (Park et al. 2013) on top of the C IV λ1549Å emission line or alternatively a strong foreground absorption feature.

In Figure 4, we visualize four dimensions of the latent space, which show the strongest dependency on the black hole mass label as determined from calculating the Pearson correlation coefficients. It is evident that no one latent dimension alone encodes the information of the black hole mass but rather the black hole masses depend on a combination of the 16 latent dimensions. Note that the latent dimensions are strongly degenerate, and hence one could in principle impose a dependency of the black hole mass onto a chosen latent dimension by modifying the prior on the latent space (Equation (12)). In practice we find that imposing a more restrictive prior does not improve the predictions.

It is important to note that in the regime where all quasars have complete spectral coverage of certain emission lines (i.e., in the here chosen data set all quasars have spectral coverage of the C IV emission line), the scatter of 0.4 dex in the predicted black hole masses shown in Figure 3 is comparable to the scatter in predictions from scaling relations (e.g., Park et al. 2017). However, these scaling relations are no longer applicable for objects where these emission lines are unobserved, masked by telluric absorption, or simply very noisy, in which case the GPLVM can still produce a reliable black hole mass estimate. Furthermore, the GPLVM allows us to include more quasars in the training step that might not have coverage of certain spectral features but would nevertheless improve modeling (see Section 7).

5.4. Predicting Unobserved or "Missing" Spectral Regions

In Figure 5, we test how well the spectral features of a new quasar spectrum that was not part of the training set can be predicted by our generative model. In each panel, we use the yellow-shaded spectral region x_* (without any information about the labels y_*) to determine the latent parameters z_* using Equation (14). With this set of latent parameters we generate the "missing" spectral regions $\tilde{x}_*$ and labels $\tilde{y}_*$ by means of Equations (19) and (20).

Generally, the predictions of held-out spectral features work extremely well as long as the spectral coverage of the input features contain sufficient information. However, the accuracy of the spectral feature prediction and the black hole mass label prediction decreases with less spectral coverage as expected, and approaches something like the mean of the training-set spectra shown in gray if the input x_* is not sufficiently informative. With only 30 quasars in each leave-one-out training set, it is perhaps surprising that these predictions are so good. The quality of these predictions indicates that quasar spectra are intrinsically very low in dimensionality given that even limited spectral coverage and a limited data set can train a model that makes good predictions (with a reduced $\chi^2 \lesssim 5$, see Figure 5) of held-out spectral features.

Note that the hyperparameters Q and β were determined via cross-validation when optimizing the black hole mass predictions (Figure 3) rather than the spectral features, and thus it is likely that the predictions for the unobserved spectral regions could be further improved by choosing a different set of values for Q and particularly for β that would be optimized with respect to the spectral predictions. Furthermore, the model is trained on a limited range of quasar spectra due to the very small number of objects in the training set, and thus larger numbers of quasars in the training set will also improve this prediction.

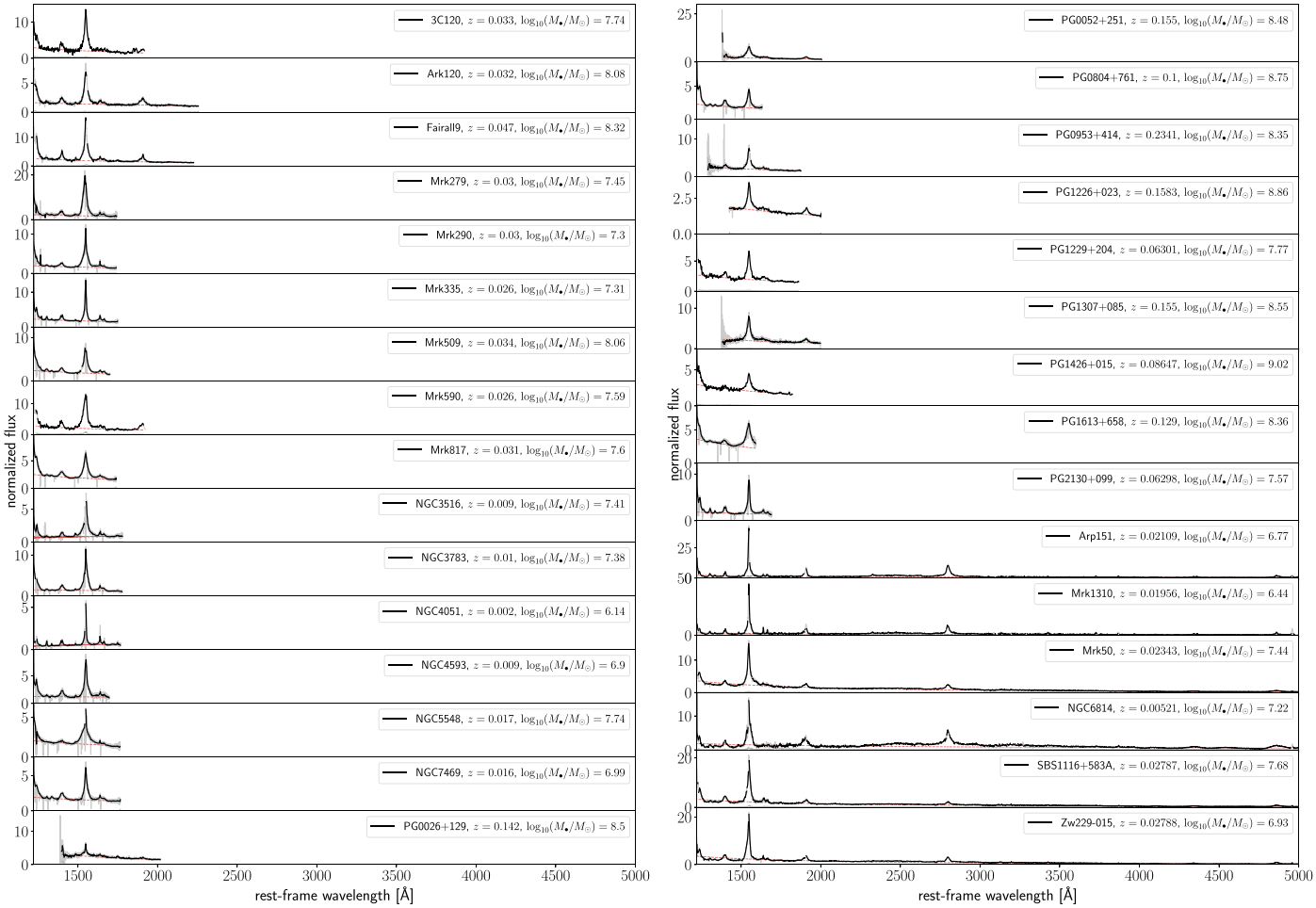

Figure 2. Quasar spectra in our data set taken with COS/HST and STIS/HST. The light gray spectra show the original data, while the black curves show the input data re-binned to a common wavelength grid and after fitting a spline and sigma-clipping to remove absorption lines. The red-dashed curve shows the power-law fit to each spectrum, which is used to normalize all spectra to unity at approximately 2500 Å.

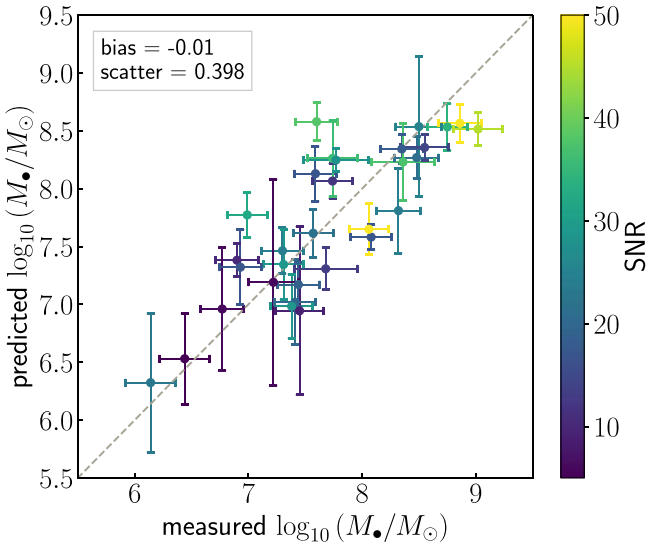

Figure 3. Cross-validation of the GPLVM with $Q = 16$ latent dimensions. Predicted compared to measured black hole masses for all quasars in our data set, colored by the SNR of the single-epoch quasar spectrum. Note that we use only two labels for the quasars here (i.e., $L = 2$), namely black hole mass and bolometric luminosity.

5.5. Spectral Dependencies on Quasar Properties

In Figure 6, we show the spectral dependencies of the quasar spectra on three labels (i.e., black hole mass, bolometric luminosity, and redshift). We randomly sample the 16-dimensional latent space and show the median of all spectra that correspond to a given black hole mass within $\Delta \log_{10}(M_\bullet/M_\odot) = 0.1$, a fixed bolometric luminosity within $\Delta \log_{10}(L_{\rm bol}/{\rm erg\ s^{-1}}) = 0.2$, and a fixed redshift within $\Delta z = 0.01$. We only show the wavelength region at $\lambda_{\rm rest} \lesssim 2000$ Å, where we see the strongest spectral differences, which is likely due to the fact that only six quasars in our current data sample cover the rest-frame optical wavelength regime (see Figure 2).

We observe a few interesting and expected trends. In the top panel of the figure showing the spectral dependencies with black hole mass, the emission lines show the expected broadening with increasing black hole mass, such as the C IV emission line, the S II+O I complex, or the N V emission line. Interestingly, the amplitude of the C IV emission line shows a stronger dependency on the black hole mass than the width of the line, which suggests that the width of the C IV emission line alone is unlikely to be a good proxy for black hole mass (see also Coatman et al. 2017).

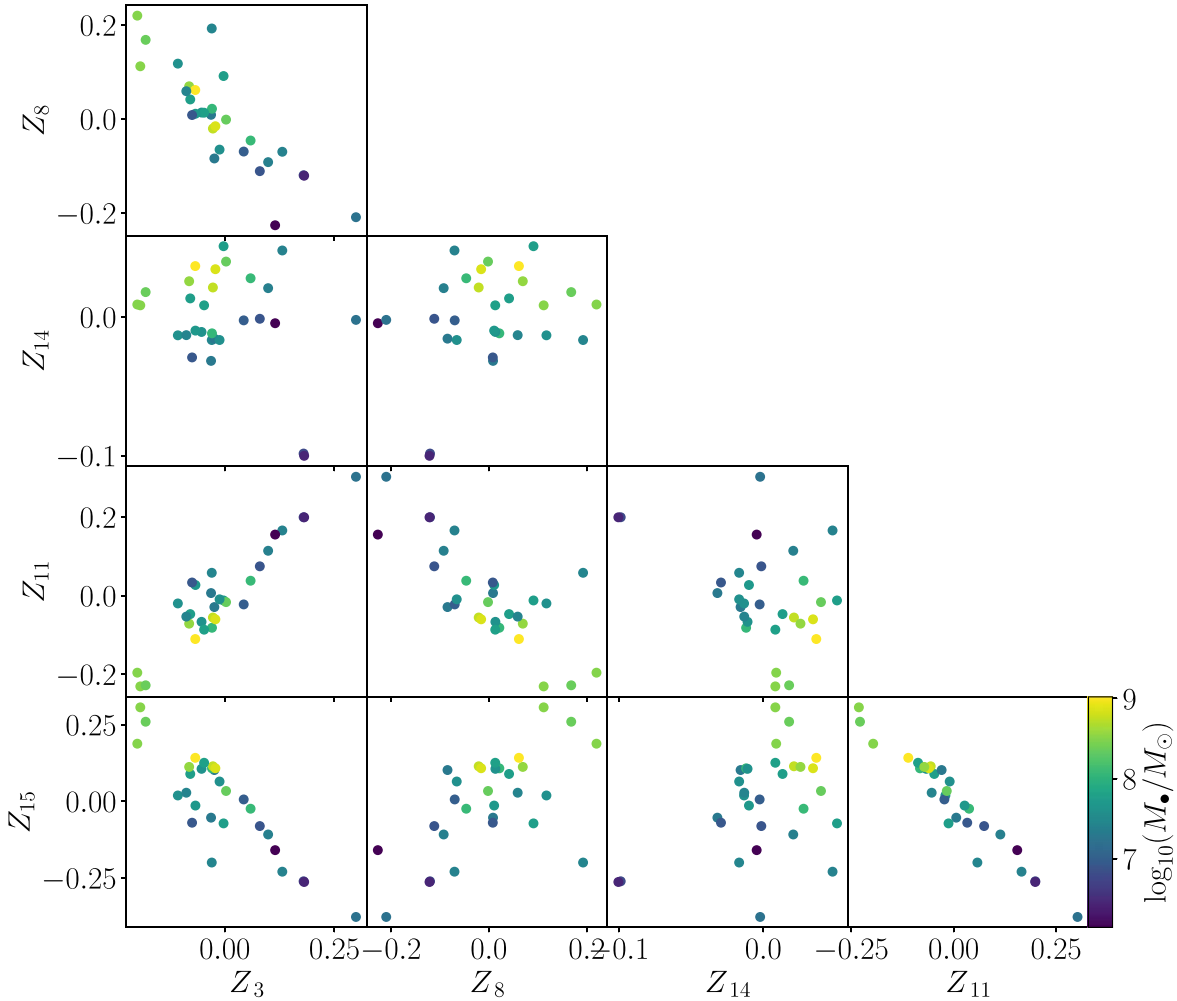

Figure 4. Visualization of the latent space. We show the locations of the quasars in five latent dimensions, which show the strongest gradient with black hole mass based on the Pearson correlation coefficient.

Furthermore, the semi-forbidden lines S III] and C III] show a less strong dependency on the black hole mass than the permitted lines. These lines generally arise from lower density gas close to the critical density—that is, $N_e \sim 3 \times 10^9\,\mathrm{cm}^{-3}$ (e.g., Osterbrock & Ferland 2006)—, which is likely located at larger radii from the black hole (e.g., AGN 1990), and thus the Doppler broadening is less apparent.

The second panel shows the trends with bolometric luminosity, where we can nicely observe the Baldwin effect (Baldwin 1977), which indicates that quasar spectra show a decreasing equivalent width of their UV and optical emission lines with increasing bolometric luminosity. Both the C IV emission line as well as some of the fainter lines such as He II, which is highlighted in the inset plot, show this effect.

We note, however, that the interpretation of this figure should be taken with caution because of the limited number of objects that is currently used for training the GPLVM. Due to the nature of our current data set, there is also a mild degeneracy between L_{bol} and $M_\bullet$ (see Figure 1). Thus, the effects on the spectra from varying these parameters are difficult to disentangle. Applying the GPLVM to a larger number of quasars that span a wide range of parameters will enable a more detailed study of the spectral dependencies with physical quasar properties in the future.

6. Current Limitations of the Model

In this work, we presented a novel generative model based on a GPLVM, which can handle data sets with heteroscedastic noise and missing data correctly. These two features are crucial for any astrophysical applications with real data. The chosen application in Section 5 represents a proof of concept and a first demonstration of the model capabilities. However, our model and its application are currently limited by two main factors. The first factor is that we only know a limited number of quasars at the moment, for which precise RM black hole mass measurements are available (see Section 7 for soon upcoming data from new surveys). The second factor is that the current implementation of the GPLVM does not use state-of-the-art tools, and hence the performance and optimization of our model is very time intensive.

Multiple future improvements on the implementation of our model are possible and are already a work in progress, such as the use of a significantly more efficient implementation of Gaussian processes, possibly with GPU acceleration; for example, `gpytorch`[10] (Gardner et al. 2018) or `tinyGP`.[11] Additionally, the use of auto-differentiation, such as that

[10] https://gpytorch.ai/
[11] https://tinygp.readthedocs.io/en/stable/

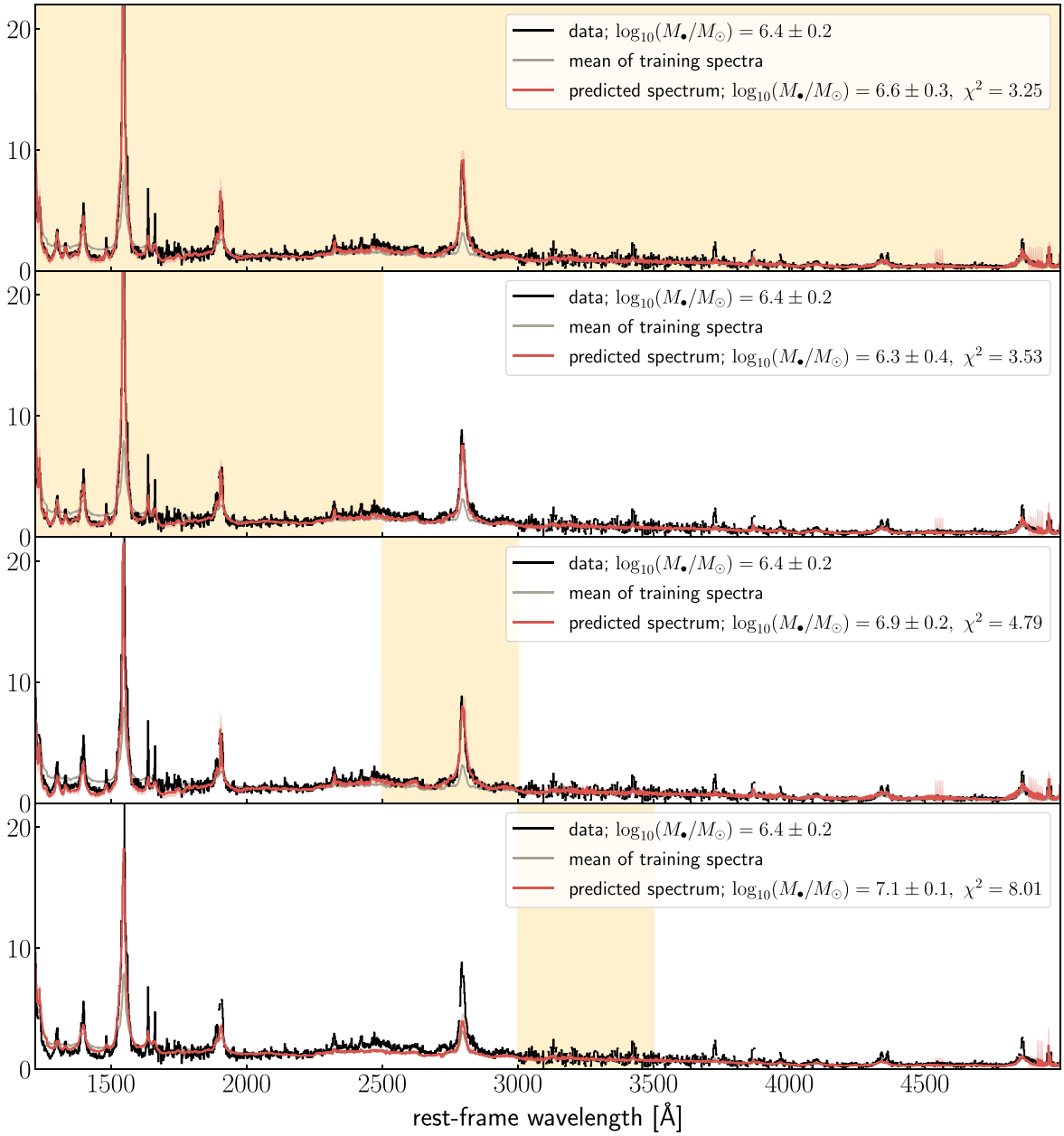

Figure 5. Predictions of missing spectral regions. In each panel, we use the spectral region indicated by the yellow-shaded regions to determine the latent representation of the new unseen quasar spectrum shown in black (see Section 3.1). The red curves show the predicted spectra derived from the set of latent parameters. As expected, the precision of the spectral prediction increases when a larger fraction of the spectrum is used to determine the latent parameters, whereas the prediction approaches the mean of the training-set spectra shown in gray if only a small fraction of the spectral coverage is provided as input.

implemented in `jax`[12], which automatically differentiates native Python code, could also reduce the complexity of the code and might even speed it up.

Nevertheless, the nature of a GPLVM is that it will always scale badly with the number of objects. Naively, the training should take computation time proportional to the cube of the number of training-set objects. Furthermore, our adaption of the algorithm to handle data with heteroscedastic uncertainties and missing data means that one cannot factorize the matrices in the likelihood functions once, which causes the algorithm to scale badly to large data sets.

However, it would be possible to use the recent 2D generalization of `celerite` (Foreman-Mackey et al. 2017) to compute the likelihoods, which would then result in a linear scaling with the number of data points (e.g., Gordon et al. 2020). Furthermore, for a better scalability of the GPLVM, one could apply minibatch training, where the training set is split into small batches that are used to update the model coefficients (e.g., Lalchand et al. 2022). While a better implementation will certainly improve the computation time and optimization of our generative model, it will still not scale well to extremely large samples. Thus, the GPLVM is a model for smaller, complex, heterogeneous data sets, where each training point was hard won.

It might, however, be considered absurd in the machine-learning or data-driven astrophysics literature to be training a model with a training set of only 31 objects! As RM projects proceed, this training set will get larger and the results will

[12] https://jax.readthedocs.io/en/latest/

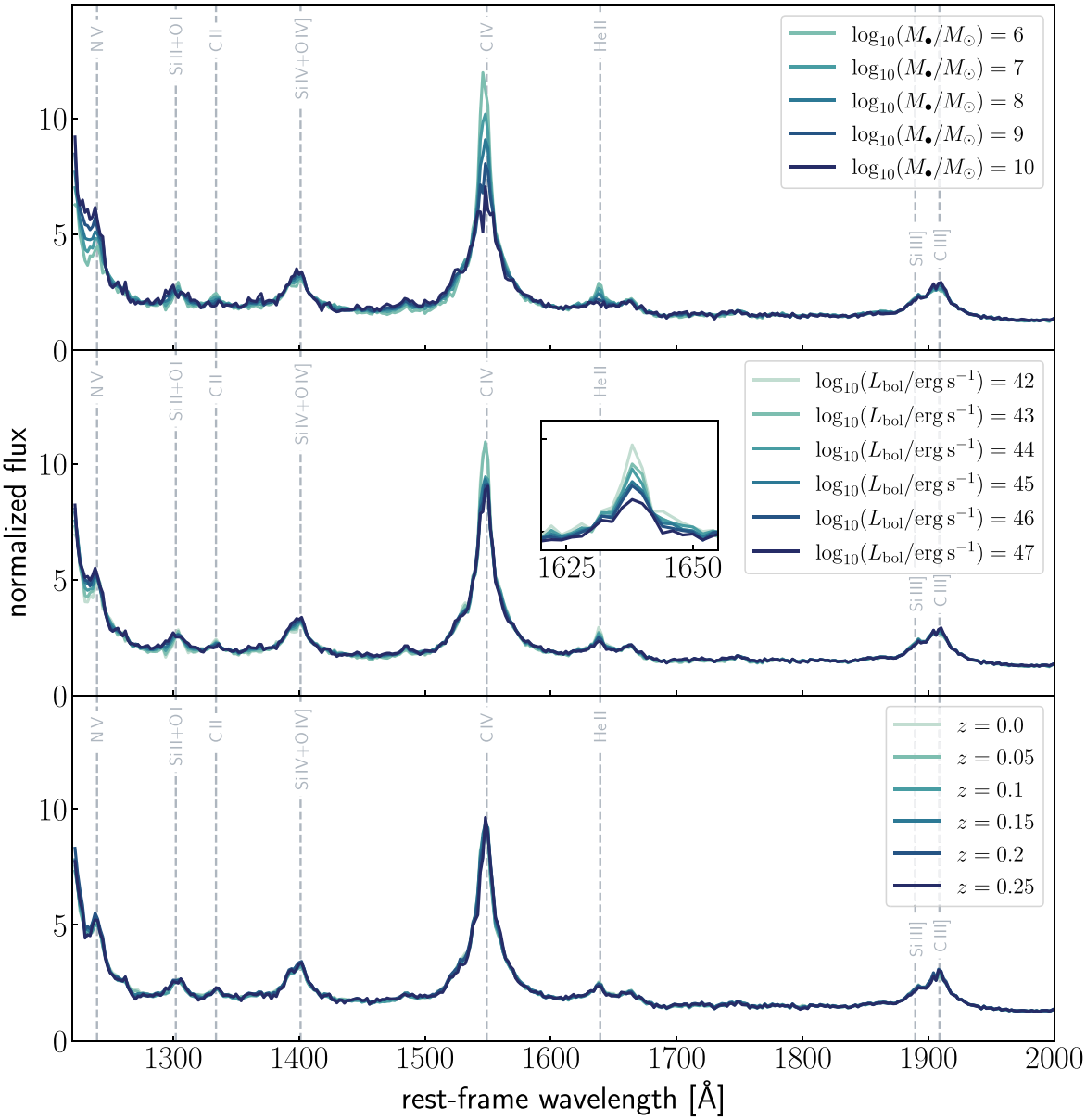

Figure 6. Spectral dependencies on various quasar properties; that is, black hole mass (top), bolometric luminosity (middle), and redshift (bottom). The inset panel shows the expected spectral changes due to the Baldwin effect.

(presumably) get better. That being said, we show that even with this tiny training set, we can predict black hole masses in held-out data about as well as might be expected given the measurement precision. This testifies to the power of this generative model, and the consistency, dimensionality, and information content of quasar spectra.

Another potential limitation of the model could be the generalisability of the model between the small subset of predominantly low-redshift quasars for which RM measurements exist and a set of high-redshift quasars for which we ultimately aim to understand their black hole masses. However, upcoming surveys such as SDSS-V (Kollmeier et al. 2017) and the Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) on the Rubin Observatory will push the RM measurements to quasars at higher redshifts, with measurements expected for quasars at redshifts up to $z \sim 4$. These measurements will help to bridge the gap between the lower-redshift quasar population that can be used as a training set for the GPLVM and the higher-redshift quasar population, which we ultimately aim to use as a testing set.

Additionally, because the GPLVM allows us to incorporate missing data or data with large measurement uncertainties in a rigorous way, we could add the spectral features of high-redshift quasars to the training set with a corresponding label for the black hole mass that is either missing completely or has a large measurement uncertainty. However, if one adds significantly more unlabeled than labeled data to the training set, then one needs to ensure that the latent space will not only learn the spectral features but also the labels. This could potentially require modifications to the model structure to give sufficient weight to the labels, and will thus be part of future work.

Finally, a substantial limitation of the model is that the model has been provided essentially no prior knowledge about either the atomic physics or the accretion physics generating the spectrum. Further improvements could likely be made with

models that include a mix of data-driven components with physically motivated components (demonstrated for instance in Leistedt & Hogg 2017).

7. Summary and Outlook

This paper presents a generative model for quasar spectra that *simultaneously* generates both the spectral features of the objects and its labels. We chose a GPLVM, which can handle heteroscedastic data sets observed with different telescopes or instruments with measurement uncertainties and missing, unobserved or unlabeled data in a principled way. Our model allows us to consistently predict quasar properties with limited spectral coverage that could be varying between different objects, as well as with noisy or only partially measured labels in the training set.

As a first application and proof of concept, we apply our model to a data set of 31 quasars with precise black hole mass measurements obtained via the RM technique and show that the model can predict the black hole mass measurements from the spectral features of an unseen quasar close to the best possible precision. Most importantly, we show that the GPLVM can obtain estimates for the black hole mass of a quasar from a limited spectral region. This has the advantage that specific emission lines (e.g., Hβ or Mg II), which might not always be observable due to atmospheric absorption for instance, are no longer required.

The scope of this first application is currently still limited for two reasons. The first reason is that there are only a very limited number of quasars known that we have precise black hole mass measurements based on the RM technique. The second reason is that the current implementation of the model can be significantly improved by using state-of-the-art algorithms, as discussed in Section 6. Both limitations will be overcome in the future: work on a more sophisticated implementation of the algorithm is ongoing, and upcoming surveys such as SDSS-V and LSST promise an increase of 2–3 orders of magnitude in precise RM measurements for quasars up to redshift $z \sim 4$ within the next decade (e.g., Kollmeier et al. 2017; Ivezić et al. 2019). These improvements will enable us to constrain precise black hole masses for quasars at all redshifts, which is crucial for understanding the co-evolution of galaxies and their central SMBHs (e.g., Volonteri 2012).

7.1. Future Applications of the GPLVM

Our generative model has many possible applications. We will briefly discuss a few of these applications in this subsection.

1. Most single-epoch black hole mass measurements are derived using the width of the Hβ emission line and scaling relations calibrated to low-redshift quasars with RM measurements (e.g., Park et al. 2013, 2017; Grier et al. 2017). However, for quasars at high redshifts of $z \gtrsim 3$, the Hβ emission line is not observable with ground-based observatories and thus the scaling relations are re-calibrated to the still observable rest-frame UV emission lines, such as C IV or Mg II (e.g., Coatman et al. 2017). For $z \gtrsim 5$ quasars, these emission lines often fall into regions of significant telluric absorption at near-IR wavelengths. Thus, even the single-epoch black hole mass scaling relations cannot be applied to these objects.

   By using the GPLVM, we can omit all calibration and scaling steps that cause additional uncertainties and possible biases in the black hole mass estimates because the generative model does not need the full spectral coverage or coverage of a specific emission line to determine the unknown labels. Thus, predicting the black hole masses from single-epoch spectra of $z \gtrsim 5$ quasars using generative models such as the GPLVM circumvents these limitations and might therefore result in more accurate predictions than conventional scaling relations.
2. As shown in Section 5.4, the generative model can be used to predict unobserved or "missing" spectral regions, which could also be useful to predict the unabsorbed continuum emission of high-redshift quasars in the Lyman-series forest. For quasars at $z \gtrsim 5$, a significant fraction of their continuum emission at wavelengths shorter than Lyα at $\lambda_{\rm rest} = 1215.67$ Å is absorbed due to the high fraction of neutral hydrogen in the surrounding IGM. However, an accurate knowledge of the quasars' unabsorbed continuum emission is essential for analyses of the neutral fraction of the IGM by means of the IGM damping wing (e.g., Simcoe et al. 2012; Davies et al. 2018a; Greig et al. 2022), measurements or the IGM opacity in the Lyα or Lyβ forests (e.g., Fan et al. 2006; Becker et al. 2015; Eilers et al. 2018, 2019; Yang et al. 2020; Bosman et al. 2022), or measurements of the quasars' proximity zone sizes (e.g., Eilers et al. 2017, 2020; Chen et al. 2022; Morey et al. 2021). To this end, many studies have attempted to predict the unabsorbed quasar emission using approaches such as PCA (e.g., Suzuki et al. 2005; Pâris et al. 2011; Davies et al. 2018b; Bosman et al. 2021), neural nets (e.g., Ďurovčíková et al. 2020; Liu & Bordoloi 2021), normalizing flows (Sun et al. 2022), or constructing composite spectra of nearest neighbors in low-redshift quasar spectra (e.g., Simcoe et al. 2012) to accurately predict the emission.

   However, an important shortcoming of all of these approaches is that they are trained using low-redshift quasars, where there is significantly less absorption from the intervening IGM and the continuum emission can be more easily be reconstructed. This approach implicitly assumes that there is no redshift evolution in the spectral shape of quasars. However, we know that this is not a good assumption because we observe differences in the composite spectra of low- and high-redshift quasars (Shen et al. 2019; Yang et al. 2021); for example, emission lines are often more blueshifted with respect to the quasars' systemic redshifts in high-redshift quasars (e.g., Meyer et al. 2019), which could lead to biases in the continuum reconstruction. The advantage of the GPLVM presented here is that it can deal with data sets with heteroscedastic noise and missing data, and thus we can include the spectra of both low- and high-redshift quasars in the training set by assuming "missing" spectral coverage bluewards of the Lyα line for the high-redshift quasar spectra, which are heavily affected by IGM absorption.
3. Several quasar properties require expensive and time consuming observations to be determined, such as measurements of the systemic redshift of a quasar. The most reliable redshift estimates are based on sub-mm emission lines, such as [C II] at 158 μm, which are the dominant cooling mechanism of the interstellar medium in the quasars' host galaxies (e.g., Carilli & Walter 2013). In contrast, broad rest-frame UV and optical emission lines are subject to strong internal motions or winds in the

BLR, and thus are often displaced from the systemic redshift (e.g., Richards et al. 2002; Meyer et al. 2019). However, obtaining sub-mm observations with the Atacama Large Millimetre Array (ALMA), for instance, is highly competitive and expensive, and thus one could attempt to infer the quasar's systemic redshifts by means of this generative model because the velocity shifts of different rest-frame UV and optical lines likely encode information about the quasar's systemic redshift (see also Fauber et al. 2020).

It is our pleasure to thank Joe Hennawi, Robert Simcoe, Hans-Walter Rix, Aaron Barth, Adrian Price-Whelan, Vidhi Lalchand, and Vincent Sitzmann for very helpful discussions. Furthermore, we are grateful to Daesong Park for sharing the HST data, as well as to Joe Hennawi for the use of the computing cluster at UCSB.

This project was developed in part at the 2017 Heidelberg Gaia Sprint, hosted by the Max-Planck-Institut für Astronomie, Heidelberg.

A.C.E. acknowledges support by NASA through the NASA Hubble Fellowship grant #HF2-51434 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555.

J.T.S. acknowledges funding through the ERC European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement No 885301).

Software: numpy (van der Walt et al. 2011), scipy (Virtanen et al. 2020), matplotlib (Hunter 2007), astropy (The Astropy Collaboration et al. 2018).

ORCID iDs

Anna–Christina Eilers https://orcid.org/0000-0003-2895-6218
David W. Hogg https://orcid.org/0000-0003-2866-9403
Bernhard Schölkopf https://orcid.org/0000-0002-8177-0925
Daniel Foreman-Mackey https://orcid.org/0000-0002-9328-5652
Frederick B. Davies https://orcid.org/0000-0003-0821-3644
Jan–Torge Schindler https://orcid.org/0000-0002-4544-8242

References

AGN 1990, Active Galactic Nuclei (Berlin: Springer)
Baldwin, J. A. 1977, ApJ, 214, 679
Barth, A. J., Bennert, V. N., Canalizo, G., et al. 2015, ApJS, 217, 26
Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, MNRAS, 447, 3402
Bentz, M. C., Peterson, B. M., Netzer, H., Pogge, R. W., & Vestergaard, M. 2009, ApJ, 697, 160
Blandford, R. D., & McKee, C. F. 1982, ApJ, 255, 419
Bosman, S. E. I., Ďurovčíková, D., Davies, F. B., & Eilers, A.-C. 2021, MNRAS, 503, 2077
Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, MNRAS, 514, 55
Carilli, C. L., & Walter, F. 2013, ARA&A, 51, 105
Chen, H., Eilers, A.-C., Bosman, S. E. I., et al. 2022, ApJ, 931, 29
Coatman, L., Hewett, P. C., Banerji, M., et al. 2017, MNRAS, 465, 2120
Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018a, ApJ, 864, 142
Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018b, ApJ, 864, 143
Ďurovčíková, D., Katz, H., Bosman, S. E. I., et al. 2020, MNRAS, 493, 4256
Eilers, A.-C., Davies, F. B., Hennawi, J. F., et al. 2017, ApJ, 840, 24
Eilers, A.-C., Hennawi, J. F., & Davies, F. B. 2018, ApJ, 867, 30
Eilers, A.-C., Hennawi, J. F., Davies, F. B., & Oñorbe, J. 2019, ApJ, 881, 23
Eilers, A.-C., Hennawi, J. F., Decarli, R., et al. 2020, ApJ, 900, 37
Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, AJ, 132, 117
Fauber, L., Ho, M.-F., Bird, S., et al. 2020, MNRAS, 498, 5227
Fausnaugh, M. M., Grier, C. J., Bentz, M. C., et al. 2017, ApJ, 840, 97
Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220
Gao, X., Wang, X., Tao, D., & Li, X. 2011, ITSMC, 41, 425
Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., & Wilson, A. G. 2018, arXiv:1809.11165
Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJL, 539, L13
Gordon, T. A., Agol, E., & Foreman-Mackey, D. 2020, AJ, 160, 240
Greig, B., Mesinger, A., Davies, F. B., et al. 2022, MNRAS, 512, 5390
Grier, C. J., Trump, J. R., Shen, Y., et al. 2017, ApJ, 851, 21
Gültekin, K., Richstone, D. O., Gebhardt, K., et al. 2009, ApJ, 698, 198
Häring, N., & Rix, H.-W. 2004, ApJL, 604, L89
Hunter, J. D. 2007, CSE, 9, 90
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv:1711.03234
Lalchand, V., Ravuri, A., & Lawrence, N. D. 2022, arXiv:2202.12979
Lawrence, N. 2003, in Advances in Neural Information Processing Systems, Vol. 16, ed. S. Thrun, L. Saul, & B. Schölkopf (Cambridge, MA: MIT Press)
Lawrence, N. 2005, JMLR, 6, 1783
Lawrence, N. D., & Moore, A. J. 2007, in Proc. 24th Int. Conf. on Machine Learning, ICML '07 (New York: ACM), 481
Leistedt, B., & Hogg, D. W. 2017, ApJ, 838, 5
Liu, B., & Bordoloi, R. 2021, MNRAS, 502, 3510
Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285
McLure, R. J., & Dunlop, J. S. 2002, MNRAS, 331, 795
Meyer, R. A., Bosman, S. E. I., & Ellis, R. S. 2019, MNRAS, 487, 3305
Morey, K. A., Eilers, A.-C., Davies, F. B., Hennawi, J. F., & Simcoe, R. A. 2021, ApJ, 921, 88
Onken, C. A., Ferrarese, L., Merritt, D., et al. 2004, ApJ, 615, 645
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei (Sausalito, CA: Univ. Science Books)
Pâris, I., Petitjean, P., Rollinde, E., et al. 2011, A&A, 530, A50
Park, D., Barth, A. J., Woo, J.-H., et al. 2017, ApJ, 839, 93
Park, D., Woo, J.-H., Denney, K. D., & Shin, J. 2013, ApJ, 770, 87
Pensabene, A., Carniani, S., Perna, M., et al. 2020, A&A, 637, A84
Peterson, B. M. 1993, PASP, 105, 247
Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning) (Cambridge, MA: The MIT Press)
Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006, ApJS, 166, 470
Richards, G. T., Vanden Berk, D. E., Reichard, T. A., et al. 2002, AJ, 124, 1
Shen, Y., Brandt, W. N., Richards, G. T., et al. 2016, ApJ, 831, 7
Shen, Y., Wu, J., Jiang, L., et al. 2019, ApJ, 873, 35
Simcoe, R. A., Sullivan, P. W., Cooksey, K. L., et al. 2012, Natur, 492, 79
Sun, Z., Ting, Y.-S., & Cai, Z. 2022, arXiv:2207.02788
Suzuki, N., Tytler, D., Kirkman, D., O'Meara, J. M., & Lubin, D. 2005, ApJ, 618, 592
The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Tipping, M. E., & Bishop, C. M. 1999, J. R. Stat. Soc. Series B Stat. Methodol., 61, 611
Titsias, M., & Lawrence, N. D. 2010, in Proc. Machine Learning Research, Vol. 9, Proc. of the Thirteenth Int. Conf. on Artificial Intelligence and Statistics, ed. Y. W. Teh & M. Titterington (Sardinia: PMLR), 844
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Vestergaard, M., & Osmer, P. S. 2009, ApJ, 699, 800
Vestergaard, M., & Peterson, B. M. 2006, ApJ, 641, 689
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Volonteri, M. 2012, Sci, 337, 544
Wang, J.-G., Dong, X.-B., Wang, T.-G., et al. 2009, ApJ, 707, 1334
Williams, C. K. I. 1998, in Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond, ed. M. I. Jordan (Dordrecht, Netherlands: Springer), 599
Woo, J.-H., Yoon, Y., Park, S., Park, D., & Kim, S. C. 2015, ApJ, 801, 38
Yang, J., Wang, F., Fan, X., et al. 2020, ApJ, 904, 26
Yang, J., Wang, F., Fan, X., et al. 2021, ApJ, 923, 262
Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. 1997, ACM Trans. Math. Softw, 23, 550