Composite Spectral Energy Distributions and Bayesian Machine Learning for Spectral Data

Ming-Feng Ho[1]

[1]Department of Physics and Astronomy, University of California, Riverside; email: mho026@ucr.edu

Xxxx. Xxx. Xxx. Xxx. 2019. AA:1–18

https://doi.org/10.1146/((no doi avaliable))

Keywords

galaxies: evolution, galaxies: high-redshift, methods: data analysis, methods: Bayesian non-parametric

Abstract

Extracting more information from photometric data can help astronomers to make intelligent decisions before taking the spectra. Multi-wavelength photometry can achieve deeper and wider areas in the space and at the same time much cheaper than spectroscopic. Composite SED method provides a new way to study high-redshift galaxies in an empirical way. By building composite SEDs from grouping multi-wavelength photometric data, we can classify different galaxy types based on the spectral features shown in the composite SEDs. This new grouping method is potentially useful for classifying rare galaxy populations with.

We will also review two Bayesian machine learning method, Gaussian process and Dirichlet process, and the usage of Bayesian machine learning for modeling the spectral data.

Contents

1. INTRODUCTION

One fundamental dilemma in observational astronomy is that: do we want to take a spectrum or take a multi-wavelength photometric observation? Taking a spectrum will give us more detailed information about the properties of an object, e.g., emission lines and absorption lines. However, the exposure time of taking a spectrum is much longer than taking a multi-wavelength photometric observation.

Using multi-wavelength photometry, on the other hand, is less expensive than taking a spectrum. The other advantage of multi-wavelength observations is that we can reach a deeper redshift and a wider spatial range. However, the trade-off of multi-wavelength observations is that we lose the resolution of the spectral features of data.

The dilemma here is similar to the problem of **exploration or exploitation** situation[1]: whether you want to focus on taking accurate observations on a small number of objects or you want to explore a wider and deeper area but lose the accuracy? Another analogy to the exploration or exploitation problem is the "Battleship" game[2]. In the Battleship game, we have to destroy our enemy's ships by guessing the locations of ships with a few trail shootings. We have to make an intelligent choice to balance exploration and exploitation to win. In typical cases, we would start by exploring the whole area; after gaining some low-resolution knowledge, we could start to focus on the plausible regions to find our enemy's hidden ships.

The game played by astronomers here is similar: should we choose to explore the space with a lower resolution by multi-wavelength photometry or exploit on a limited number of objects for taking expensive spectra? To make an intelligent decision, we need to know more. Since the budget is limited and the number of spectra we can take is also limited, gaining more information from multi-wavelength photometric data before taking spectra is essential.

EAZY: : **E**azy and **A**ccurate **Z**photo from **Y**ale

sed: : Spectral Energy Distribution

Astronomers developed techniques including UVJ color-color diagram and SED fitting to gain more information from photometric data in order to make an optimal decision. The color-color diagram classifies galaxies into quiescent or star-forming types; SED fitting estimates the properties of galaxies and redshifts (e.g., EAZY (Brammer, van Dokkum & Coppi 2008)) with the help of modeling of spectral energy distributions (SEDs) and synthetic templates built by matrix factorization (Blanton & Roweis 2007).

[1] A dilemma situation in the active (machine) learning problems.

[2] `https://en.wikipedia.org/wiki/Battleship_(game)`

One thing we can ask is: can we gain more knowledge about our targets beyond the UVJ diagram and SED fitting? Composite SED technique (Kriek et al. 2011; Forrest et al. 2018) provides an empirical way to cluster multi-wavelength targets together based on the shapes of SED fittings. With the advance of medium-band photometry (See Fig1), we can achieve a finer sampling on the shapes of SEDs and gain a better set of composite SEDs.

By examining composite SEDs, the rare populations, which are not easy to be identified in the UVJ diagram, could be classified based on the shapes of SEDs. If we gain more knowledge about our targets by using composite SEDs, astronomers can make a better decision on which target is much more interesting to take a spectrum. Usually, the marginal cases would be selected to be studied in more details[3].

2. GALAXY EVOLUTION IN TERMS OF COMPOSITE SPECTRAL ENERGY DISTRIBUTIONS (seds)

The development of composite SED method came from the improvement of medium-band photometry and the advance of SED fittings. In this section, we will first review the medium-band photometry technique. Next, we will talk about the grouping method implemented in Kriek et al. (2011); Forrest et al. (2018). In the final part of this section, we will examine the classification of galaxy types via composite SEDs using the spectral features such as H_α and $D(4000)$, and then we will talk about how these galaxy types are distributed on the UVJ diagram.

$D(4000)$: : Balmer/4 000 break

EW_{H_α}: : the equivalent width of the H_α emission line

2.1. Medium-Band Photometry at Near-IR

The choice of photometric bands is crucial for accurate redshift estimations. And without accurate redshift estimations, astronomers will be unable to achieve accurate rest-frame colors.

Past deep optical surveys such as Classifying Object by Medium-Band Observations-17 (COMBO-17, Wolf, C. et al. (2003)) and Cosmic Evolution Survey (COSMOS, Scoville et al. (2007)) achieved accurate photometric redshift accuracies with $\sigma_{\Delta z/(1+z_{spec})} \sim 1\%$ using medium-band photometry. These surveys achieve a large sample size for $z \sim 1$; however, if we want to study high-redshift galaxies up to $z \sim 3$, we need to shift our filters to a higher wavelength region.

The prominent spectral feature such as Balmer/4 000 break is shifted to near-IR region at $z \sim 1.5$. To better sample the spectral features of high-redshift galaxies, the NEWFIRM Medium-Band Survey (NMBS, Whitaker et al. (2011)) and the FourStar galaxy evolution survey (ZFOURGE, Straatman et al. (2016)) placed multiple medium bands at near-IR regions (See Fig 1). Fig 2 shows, with medium-band filters in near-IR, `FourStar` achieved a better constraint on the posterior of the photometric redshift.

2.2. Composite Spectral Energy Distributions (seds)

Composite SED method gives us a new way to classify rare galaxy types using empirical SEDs. By grouping galaxies with similar SED shapes, we can increase the signal to noise and

[3] A similar technique in active learning is called "uncertainty sampling". In uncertainty sampling, we choose to study the most uncertain targets. By focusing on the boundary targets, we can learn faster (achieve a faster convergence). See Garnett (2018).

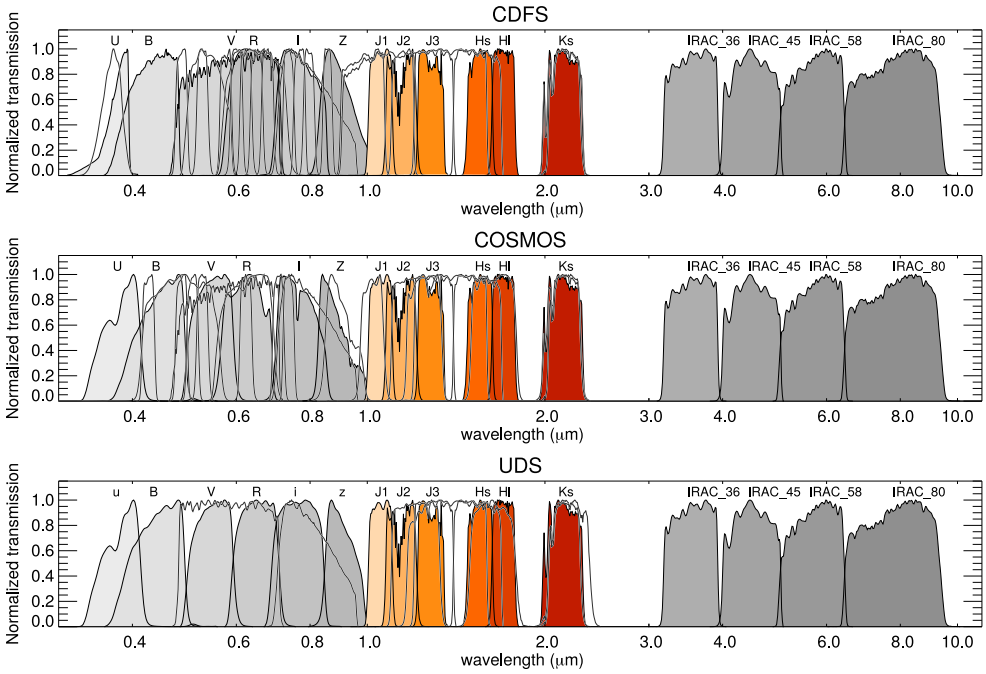

Figure 1

`Fourstar` medium bands (J1, J2, J3, Hs, Hl, Ks) in Straatman et al. (2016) (highlighted as red and yellow). The choice of medium bands would increase the resolution of sampling on the SED.

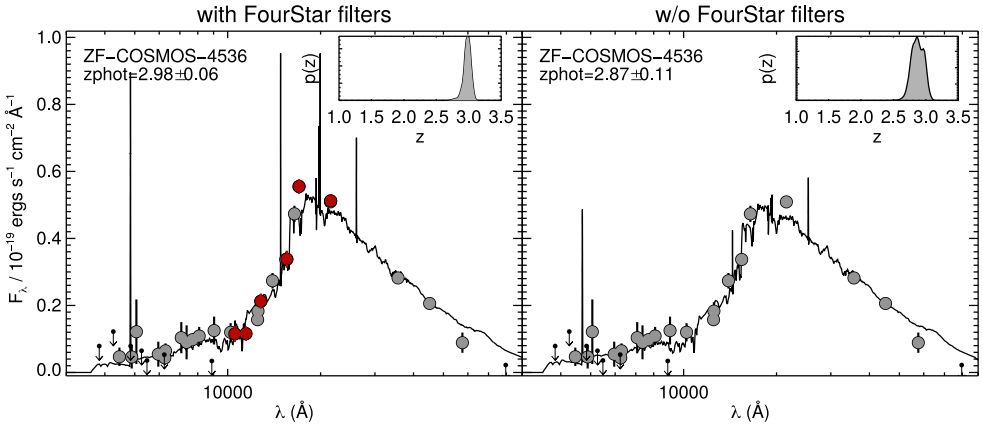

Figure 2

Sampling SED with or without `Fourstar` medium-band filters. With the help of `Fourstar` filters (red), the posterior distribution of photometric redshift (with SED fitting) is way more accurate.

build the empirical SEDs of different galaxy types.

The general grouping method is described in Fig 3 and Algorithm 1. Three galaxy types identified in Kriek et al. (2011) are shown in Fig 5 and six galaxy types identified in Forrest et al. (2018) are shown in Fig 6. By using a larger dataset ($\sim$ 7 000 galaxies) than NMBS ($\sim$ 3 500 galaxies), ZFOURGE were able to pick up rare galaxy types such as **transitional galaxies** and **post-starbursts**.

2.2.1. Grouping Method. The grouping approach which Kriek et al. (2011) and Forrest2018 took to construct composite SEDs is to compute the similarities (a root-mean-squared error metric) between different multi-wavelength photometric data at rest-frame.

Before we compute the similarity, we have to move the observed photometry back to rest-frame. The z we used to de-redshift (move to rest-frame) is estimated from SED template fitting, EAZY (Brammer, van Dokkum & Coppi 2008). After we de-redshift the observed photometry back to rest-frame, another problem arises: though we have our observed photometry data in the same wavelengths at observed wavelengths, they do not have the same rest-frame wavelengths. How can we compare two galaxies with fluxes at different rest-frame wavelengths?

To resolve the problem, 22 artificial rest-frame filters were designed. After we fit with the SED templates, we de-redshift the fitted SED back to rest-frame and bin the SED with 22 artificial filters. Therefore, the photometric data from different redshifts are comparable now because they fall within the same 22 filters.

After binning with the rest-frame filters, we compute the similarity metric,

$$b_{12} = \sqrt{\frac{\sum (f_\lambda^{rf1} - a_{12} f_\lambda^{rf2})^2}{\sum (f_\lambda^{rf1})^2}} \tag{1.}$$

with the scale factor a_{12} between different galaxies,

$$a_{12} = \frac{\sum f_\lambda^{rf1} f_\lambda^{rf2}}{\sum (f_\lambda^{rf2})^2}. \tag{2.}$$

b_{12}: : the similarity between the SED shapes of two galaxies

a_{12}: : the scale factor between the average fluxes of two galaxies

Basically, the scale factor a_{12} helps us to shift the observed fluxes to the same average flux, and the similarity metric b_{12} gives us the deviations between two different SED shapes.

In Forrest et al. (2018), they defined a threshold ($b_{12} < 0.05$) to group similar SED shapes. We first search the galaxies with largest number of analogues (galaxies with $b_{12} < 0.05$). Next, we group them together and remove these galaxies from our samples. We iterate this process until there is no galaxy has more than 19 analogues.

2.2.2. Classification of Galaxy Types. The classification of composite SEDs is done by visually inspecting the following two spectral features,

- EW_{H_α} : H_α equivalent width is an indicator for star-forming activities. Galaxies with a wider EW_{H_α} in their composite SED will have a higher chance to be star-forming galaxies. In Forrest et al. (2018), EW_{H_α} is around 100 Å for a typical star-forming galaxy. The correlation between star formation rate and the EW_{H_α} is shown in the right panel of the Fig 7.
- $D(4000)$: the Balmer/4 000 Å break indicates the age of a galaxy. Older galaxies generally have higher $D(4000)$s.

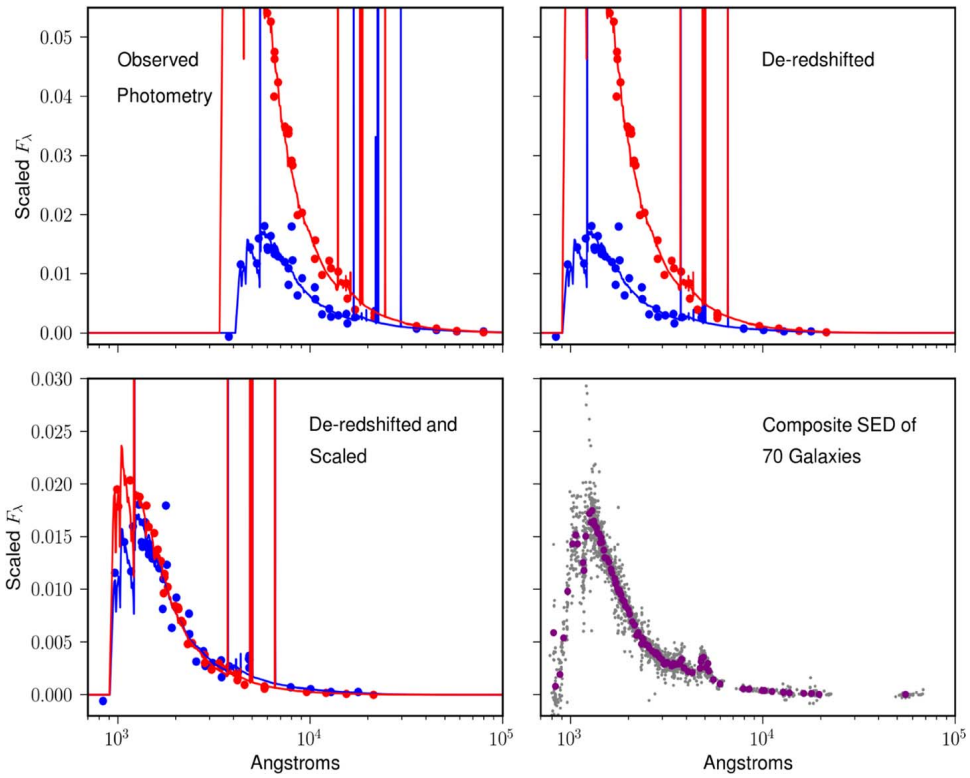

Figure 3

Grouping method for composite SED in Forrest et al. (2018). First, we take the observed multi-wavelength photometry. Second, de-redshift the wavelengths to rest-frame. Next, scale the flux using a_{12} (see Algorithm 1). In the final panel, we get the composite SED based on the similarity metric $b_{12} < 0.05$.

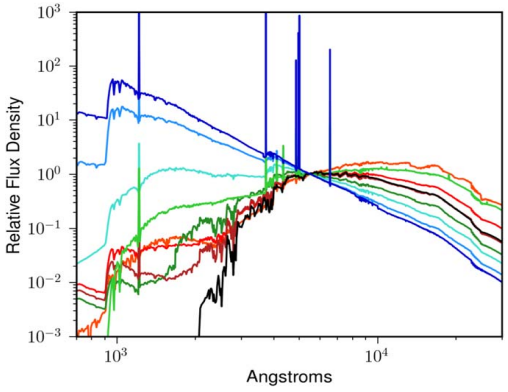

Figure 4

EAZY (Brammer, van Dokkum & Coppi 2008) templates were used in fitting the SED of galaxies in Forrest et al. (2018).

Data: Multi-wavelength photometry of galaxies within some redshift range
Result: Composite SEDs of different galaxy types
// Grab your data: fluxes and wavelengths
$\mathbf{F}, \Lambda \leftarrow$ Multi-photometry of galaxies
// Bin: 22 rest-frame filters with equal widths between 1 226 Å$< \lambda <$ 49 580 Å
$Bin(\lambda) = \text{map} : \lambda \rightarrow \lambda[\log_{10} 1226 : \frac{1}{22}(\log_{10} 49580 - \log_{10} 1226)i]$
// Loop: modify each galaxy multi-wavelength flux
for $\mathbf{f}$ *in* $\mathbf{F}$, λ *in* $\mathbf{\Lambda}$ **do**
  // Fit: with EAZY to get continous flux function
  $f_{\text{SED}}, z_{photo} \leftarrow EAZY(\mathbf{d})$
  // De-redshift: using photo-z
  $\lambda_{rest} \leftarrow \lambda/(1 + z_{photo})$
  // Bin: update discrete flux vector with rest-frame filters
  $\mathbf{F}[i, :] \leftarrow f_{\text{SED}}[Bin(\lambda_{rest})]$
end
// Similarity: use squared error to calculate similarity
$\mathbf{b} \leftarrow \text{zeros}(N, N)$
for $\mathbf{f}_i$ *in* $\mathbf{F}$ **do**
  for $\mathbf{f}_j$ *in* $\mathbf{F}$ *if* $j > i$ **do**
    // calculate scale factor; $\circ$: Hadamard (elementwise) product
    $a_{12} \leftarrow \text{sum}(\mathbf{f}_i \circ \mathbf{f}_j)/\text{sum}(\mathbf{f}_j^2)$
    // calculate similarity
    $\mathbf{b}_{ij} \leftarrow (\text{sum}((\mathbf{f}_i - a_{12}\mathbf{f}_j)^2)/\text{sum}(\mathbf{f}_i^2))^{1/2}$
  end
end
// Get the analogues from samples with b < 0.05
$\text{ind} \leftarrow \text{argsort}(\text{sum}(\mathbf{b} < 0.05).axis(1))$
$Composite_{\text{SED}} \leftarrow \text{list}()$
for i *in* ind **do**
  // find analogues
  $i_a \leftarrow find(\text{b}_i < 0.05)$
  if $\text{length}(i_a + i) < 19$ **then**
    break
  end
  $Composite_{\text{SED}}.\text{append}(\ [i_a + i]\ )$
  // remove analogues from our samples
  $ind.\text{remove}(\ [i_a + i]\ )$
end

Algorithm 1: Grouping galaxies together based on SED shape

Combine with these two features, Forrest et al. (2018) classified five galaxy types (6 if we include dusty star-forming galaxies). Fig 8 and Fig 7 show the correlation between Balmer/4 000 break and EW_{H_α}. In Fig 8, from top left to bottom right corner are emission line galaxies (purple), star-forming galaxies (blue), transitional galaxies (green), post-starbursts (yellow), and quiescent galaxies (red).

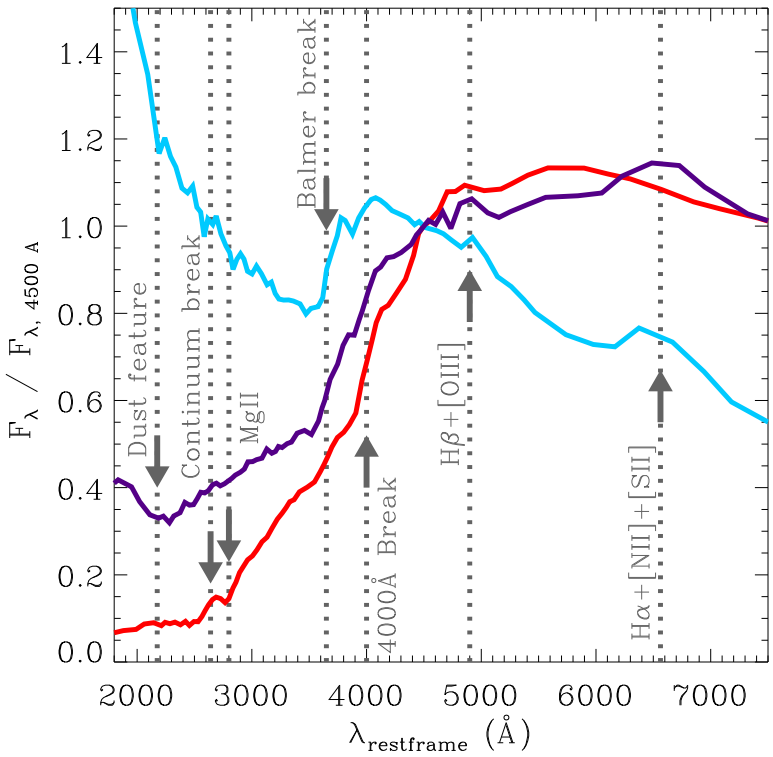

Figure 5

Three types of galaxies (red: quiescent; purple: dusty star-forming; blue: star-forming) identified in Kriek et al. (2011) using composite SED. Several spectroscopic features are labeled in the plot.

Fig 9 shows how different types of galaxies classified by composite SEDs are distributed on the UVJ color-color diagram. It is straightforward to see the composite SEDs successfully classified the quiescent galaxies (red) and star-forming galaxies (blue). Moreover, we additionally classified three more galaxy types on the color-color diagram: transitional galaxies (green), post-starbursts (yellow), and emission line galaxies (purple). Also, these rare populations are not easy to be categorized by using the UVJ color-color diagram solely.

In Forrest et al. (2018), they mentioned about the loss of diversity in the higher redshift regions. Their conclusion was that there might have less quiescent galaxies and more post-starbursts in the high-redshift regions.

ELG: : emission line galaxies

SFG: : star-forming galaxies

(D)SFG: : dusty star-forming galaixies

TG: : transitional galaxies

PSB: : post-starbursts

QG: : quiescent galaxies

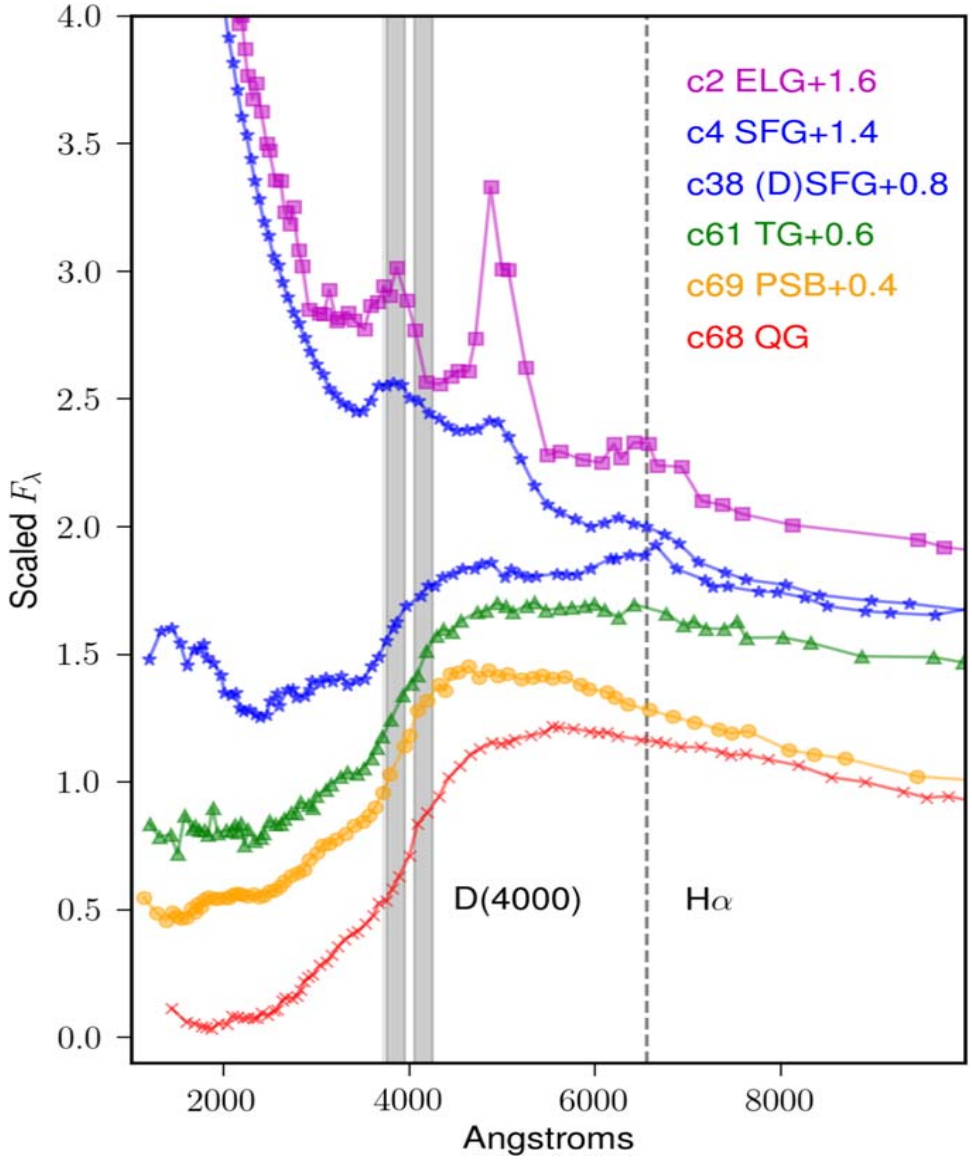

Figure 6

Six galaxy types identified by visual inspecting on two features: 4 000 break $D(4000)$ (an indicator of age) and H_α equivalent width (a probe of star formation activities).

SUMMARY POINTS

1. Composite SED is a technique to construct empirical SEDs for a given group of galaxy type and increase the signal to noise for the spectral features.
2. Spectral features such as EW_{H_α} and $D(4000)$ are used to distinguish different types

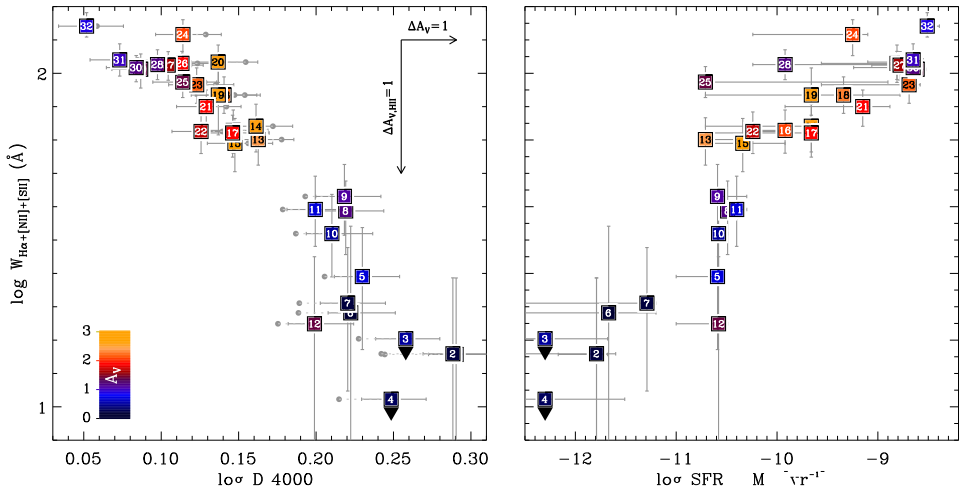

Figure 7

Left panel: correlation between H_α equivalent width and $D(4000)$; **Right panel**: correlation between H_α equivalent width and star formation rate.

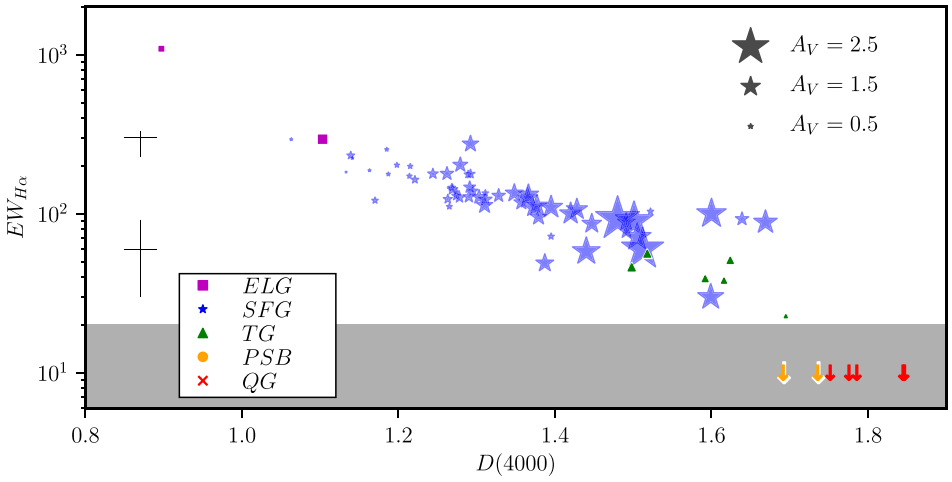

Figure 8

The correlation between H_α equivalent width and $D(4000)$ in Forrest et al. (2018). The special thing in this figure is that it labeled the types of galaxies: magenta for emission line galaxies, blue for star-forming galaxies, green for transitional galaxies (because they are in green valley), yellow for post starburst, and red for quiescent galaxies.

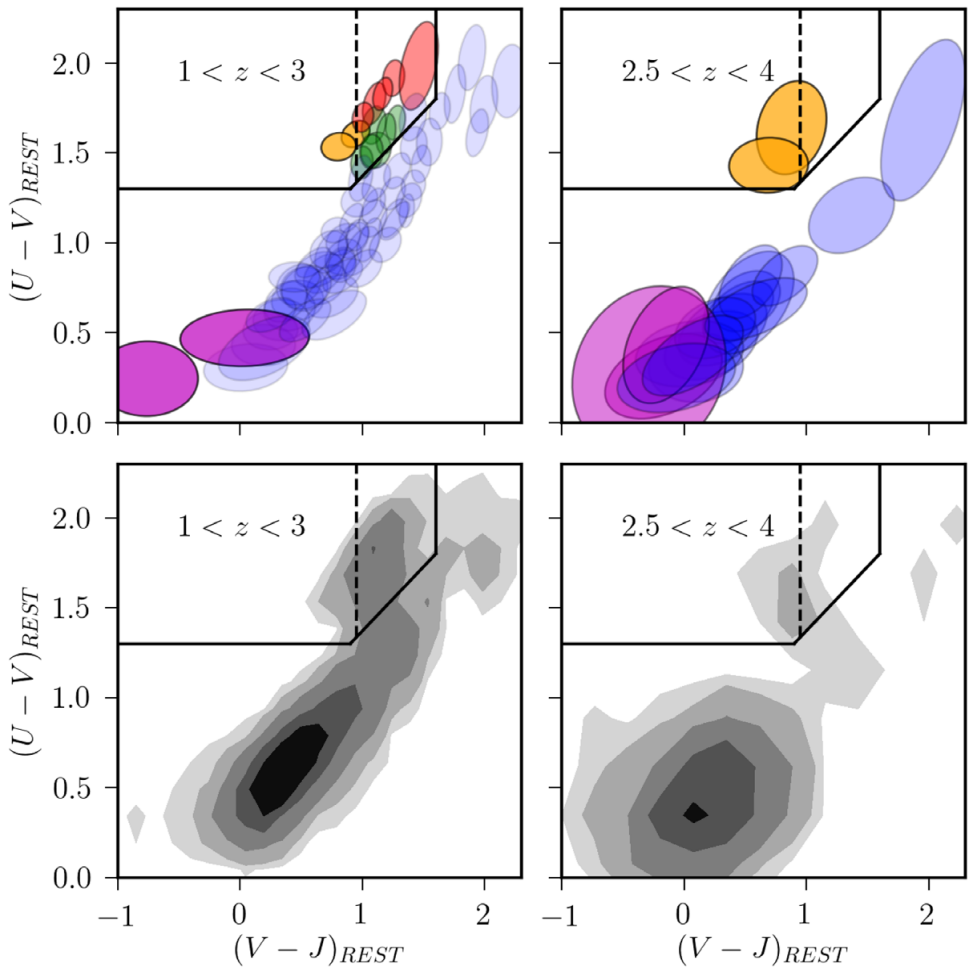

Figure 9

Color-color diagram in Forrest et al. (2018). **Top**: Different colors represent different galaxy types classified by composite SEDs. A special feature of classification using composite SED is that we can track the population of marginal (or rare) types of galaxies on the diagram, e.g., the transitional galaxies (green) and post-starbursts (yellow). **Bottom**: The distribution of number of analogue galaxies.

of galaxies.

3. Composite SEDs are able to identify rare population of galaxies, such as transitional galaxies and post-starbursts.

3. BAYESIAN MACHINE LEARNING

Bayesian machine learning is a branch of machine learning which aims to solve machine learning problems from a Bayesian perspective. Instead of optimizing the parameters from data using an empirical loss function (e.g., a least-squared function), Bayesian methods build generative models to randomly sample data from parameters and try to maximize the likelihood between observed data and hidden parameters (Barber 2012).

The difference between Bayesian statistics and Bayesian "machine learning" is that Bayesian "machine learning" is trying to approximate *non-linear* functions (Bishop & Tipping 2003). After the publish of Rasmussen & Williams (2005), learning unknown complicated functions from observed data using *Gaussian processes* (GP) became popular.

GP: : *Gaussian Processes*

$\sim$ Normal: : Sampling from a Gaussian (normal) distribution

Bayesian ML: : Bayesian Machine Learning

K: : Covariance function (matrix)

μ: : Mean function

3.1. Modeling Spectral Data using Gaussian Processes (gp)

A *Gaussian process* is a collection of random variables with any finite subset of these random variables to be a joint Gaussian distribution (Rasmussen & Williams 2005). GP could be a powerful tool to model any functional data (continuous data) in a non-parametric way. By non-parametric, it means we use infinite many parameters to describe our function (Gelman et al. 2014). GP could be treated as a random function (or a stochastic process) which draws samples from a n-dimensional distribution,

$$\mu(x_1), ..., \mu(x_n) \sim Normal((m(x_1), ..., m(x_n)), K(x_1, ..., x_n)). \tag{3.}$$

The construction of a GP could be considered as finding the mean function ($m(\vec{x})$) and a suitable covariance function ($K(\vec{x}, \vec{x}')$). In normal cases, a zero mean is usually used as a prior for GP regressions. For $K(\vec{x}, \vec{x}')$, pre-defined covariance functions (e.g., a squared exponential function $\exp(\frac{-r^2}{2\ell^2})$) are often been implemented. However, the usage of GP in modeling functional data will also be restricted by the intrinsic properties of the covariance functions. Learning a suitable covariance function is the most crucial part of machine learning in GP.

Finding a suitable choice of covariance often reflects our interpretations of the characteristics of our data (Rasmussen & Williams 2005). For example, the usage of the squared potential function $\exp(\frac{-r^2}{2\ell^2})$ implies the assumption that we believe each point on the function would have less impact to each other if they are far away on the functional space. Therefore, we need a special kind of covariance function to suit our purpose of modeling spectral data.

Garnett et al. (2017) took a machine learning approach to learn the covariance function, with a wavelength range from Ly_∞ to Lyα, from training data (quasar spectra). The optimization choice was to firstly decompose covariance matrix with (Garnett, Ho & Schneider 2015),

$$\mathbf{K} = \mathbf{M}\mathbf{M}^T, \tag{4.}$$

and then use the first 10 principle components of the flux of quasar spectra, $\mathbf{Y}$, to constitute the matrix $\mathbf{M}$. The optimization was done by maximizing the log-likelihood, $\mathcal{L}$, of the data by given $\mathbf{M}$ and absorption noise ω,

$$\mathcal{L}(\mathbf{M}, \omega) = \log p(\mathbf{Y} \mid \lambda, \mu, \mathbf{M}, \mathbf{N}, \omega, z_{\mathrm{qso}}, Model). \tag{5.}$$

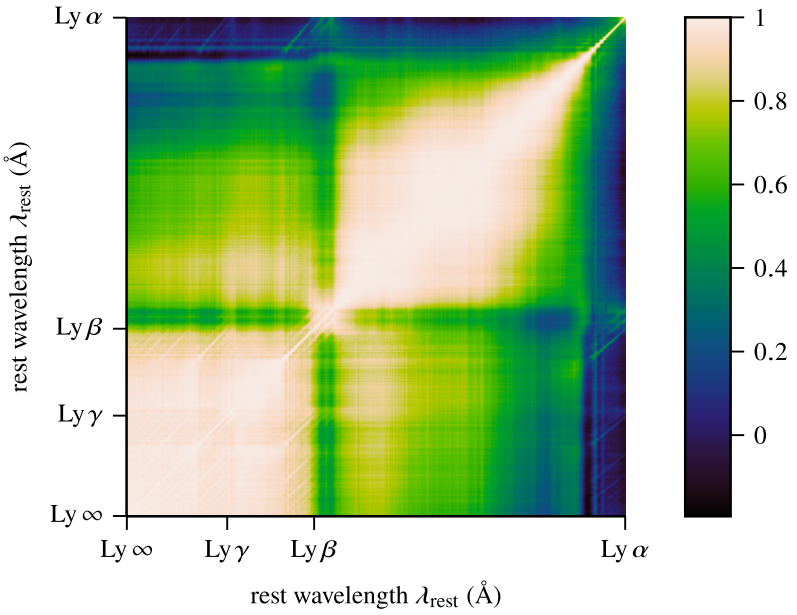

Figure 10

Covariance function for quasar spectra in Garnett et al. (2017)

The goal of optimizing above function is to find the optimal covariance matrix, $\mathbf{M}$, and absorption parameter, ω, with some given conditions. Those conditions are a given mean vector μ, the noise on the spectra $\mathbf{N}$, the redshift of the QSO, and with a given model. In a perspective of generative modeling, optimizing the data likelihood implies we are trying to find a covariance matrix to generate our spectral data better.

The covariance matrix built by Garnett et al. (2017) is shown in Fig 10. The scale in Fig 10 represents the strength of correlations between pairs of rest-frame wavelengths on the QSO spectra. The features of the Lyman series are distinct. The off-diagonal terms demonstrate the correlations of pairs of corresponding emission lines.

The mean function of GP modeling in Garnett et al. (2017) can be simply obtained by stacking the training spectra,

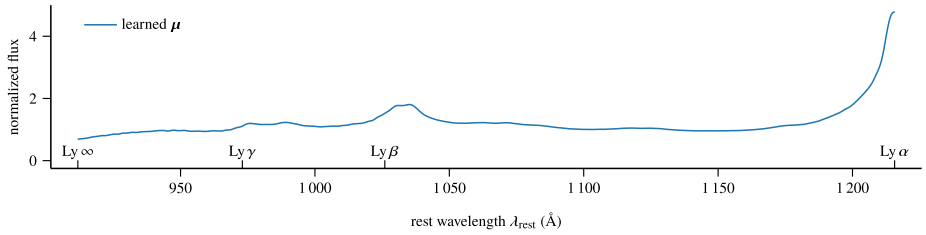

Figure 11

Mean function for quasar spectra in Garnett et al. (2017)

$$\mu_j = \text{median}(y_{ij}), \tag{6.}$$

where y_{ij} are the fluxes for spectrum. Fig 11 shows the mean function with a range from Ly$_\infty$ to Lyα. The features of emission lines on the Lyman series is also quite obvious in the Fig 11.

Generally, the GP model for spectral data could be described as

$$p(\mathbf{y} \mid \lambda, \mathbf{v}, \omega, z, Model) = Normal(\mathbf{y}; \mu, \mathbf{K} + \mathbf{\Omega} + \mathbf{V}), \tag{7.}$$

where $\mathbf{y}$ is the observed flux of the spectrum, λ is the spectroscopic grids we chose to bin the flux, $\mathbf{v}$ is the instrumental noise given by the observed data (they used SDSS QSO catalogue (Pâris, I. et al. 2012) in Garnett et al. (2017)), z is the redshift dependence of the GP model, and ω is the absorption redshift dependence (it was used to model Lyα forest in Garnett et al. (2017)).

The beauty of generative modeling the spectrum using GP framework is that we can fully control the modeling of instrumental noise and redshift dependence uncertainties. Besides, the whole framework is transparent and flexible, which implies it is interpretable and future improvements are achievable.

SUMMARY POINTS

1. *Gaussian Processes.* A flexible Bayesian non-parametric framework which allows us to model any function.
2. Learned covariance matrix $\mathbf{K}$. To model spectral data, we can optimize the covariance matrix using training data to acquire customized covariance function of any given range of wavelengths.

3.2. Dirichlet Processes

The *Dirichlet process* (Teh et al. 2006) is a Bayesian non-parametric method to model infinite mixture models and also could be used for clustering. We haven't seen any application of *Dirichlet processes* (DP) in spctral data clustering. A recent astronomical application of DP is building a mixture model for binary neutron stars in the gravitational-wave data

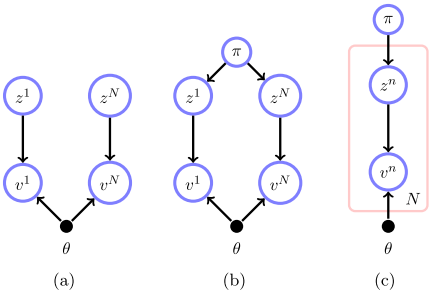

Figure 20.12: **(a)**: A generic mixture model for data $v^{1:N}$. Each z^n indicates the cluster of each datapoint. θ is a set of parameters and $z^n = k$ selects parameter θ^k for datapoint v^n. **(b)**: For a potentially large number of clusters one way to control complexity is to constrain the joint indicator distribution. **(c)**: Plate notation of (b).

Figure 12

The probabilistic graphical model of DP in plate notation in Barber (2012). The symbols here: z is the indicator and $p(z^n)$ implies the prior over n^{th} cluster; v is data; θ is used to describe the parameters on the cluster model (the shared model) while z could be treated as the hidden variable used to described component models (the individual models); π here is the parameter of a Dirichlet distribution, which is used to describe distributions.

(Del Pozzo et al. 2018). Since we expect the application of DP in the clustering of spectral data is achievable in future; we decided to roughly review the basic concept of DP here.

The DP is a stochastic process which is often used to model mixture models. Each random sample in DP is itself a distribution, which means we can treat random samples of a DP as cluster centers of a mixture model. Similar to GP, a finite subset of a DP could be described by Dirichlet distributions.

To model the data v using the Dirichlet mixture model, the clustering memberships could be described by the following conditional probability (Barber 2012),

$$p(v^{1:N} \mid \theta) = \sum_{z^{1:N}} p(v^{1:N} \mid z^{1:N}, \theta) p(z^{1:N}), \tag{8.}$$

where $p(z^{1:N})$ gives the priors over each cluster and θ describe the parameters of the cluster model. The choice of priors over clusters $p(z^{1:N}$ is crucial.

3.2.1. Chinese Restaurant Process. The clustering property of DP is hard to understand simply via mathematical forms above. However, there's an intuitive metaphor described in Teh et al. (2006) to mimic the process of drawing samples from a DP using a real-life example: *Chinese Restaurant process* (CRP).

The name of CRP was developed in the 1980s. CRP is a process to describe the distribution over partitions.

Now, imagine this: there is an infinite number of round tables in a Chinese restaurant, and there is also an infinite number of seats in each round table. Each customer will come to

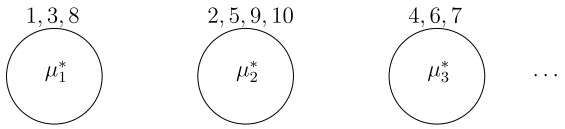

Figure 13

An illustration of CRP in Blei (2007).

the restaurant one by one (just like sampling). Let us assume the first customer choose the first table. Which table would next person choose? If every customer are friendly[4], the next customer may intend to choose the round table which has more people there. Therefore, the probability of choosing k^{th} table could be proportional to the number of people n_k in that round table. Nevertheless, she/he may also have a small amount of probability to choose a new empty table to sit depends on her/his mood of that day.

The mathematical form of CRP can be described as

$$\theta_{n+1} \mid \theta_1 ..., \theta_n \sim \frac{1}{\alpha + n} \left(\alpha H + \sum_{k=1}^{m} n_k \delta_{\theta_k^*} \right), \tag{9.}$$

where α is a constant and $\delta_{\theta_k^*} = 1$ only when θ_k^* is selected. The above equation is equivalent to CRP in this way:

1. The first customer ($n = 0$): θ_1 partition would be selected from a smooth distribution H.
2. The second customer ($n = 1$): $\theta_2 \mid \theta_1$ partition would be selected from either $\frac{\alpha H}{\alpha+1}$ (a new table) or $\frac{1}{\alpha+1}$ (the table with customer 1).
3. The $(n+1)^{th}$ customer: $\theta_{n+1} \mid \theta_n, ..., \theta_1$ would be selected from either $\frac{\alpha H}{n+\alpha}$ or $\frac{\sum_{k=1}^{m} n_k \delta_{\theta_k^*}}{\alpha+n}$.

In the above CRP scheme, the number of clusters will growth in logarithmic scale. See Fig 13 for a cartoon version of CRP. Circles represent the round tables we have in the Chinese restaurant and the numbers represent the customers. For the scheme of the 10-customer situation in Fig 13, the joint probability is

$$Pr(z_1, ..., z_{10}) = \frac{\alpha}{\alpha} \frac{\alpha}{1+\alpha} \frac{1}{2+\alpha} \frac{\alpha}{3+\alpha} \frac{1}{4+\alpha} \frac{1}{5+\alpha} \frac{2}{6+\alpha} \frac{2}{7+\alpha} \frac{2}{8+\alpha} \frac{3}{9+\alpha}. \tag{10.}$$

FUTURE ISSUES

1. **Classification of galaxy types** : the current way to label galaxy types in the composite SEDs is to visually inspect the spectral features such as the equivalent width of H_α and the Balmer/4 000 break. It is the curse of every unsupervised machine learning problems because you can not avoid to use human eyes to label

[4]There are some literature discussed the unfriendly situations.

the data. The easiest way is to use Forrest et al. (2018) as a training set, and train a supervised classifier on labeled data and use that classifier to predict the labels in the new dataset.

2. **Building dp on top of gp** : one way to do clustering functional data is to build DP on top of GP. We can first sample the partitions of cluster centers from DP, and use them as priors for the GP. One issue is that if we want to learn the covariance function and the mean function from the data, we need to have a labeled training set with a finite number of classified types of galaxies.]In this way, DP would be unable to have an infinite number of cluster centers.

ACKNOWLEDGMENTS

Thank the instructor for giving us this chance to practice writing a really long paper. It's fun because I took this chance to build my own LaTeXenvironment on my new Linux laptop. Thank Ben for the helpful discussions on the topic of composite SED. Thank Prof. Bird and Prof. Garnett for helping me to learn Bayesian machine learning.

LITERATURE CITED

Barber D. 2012. *Bayesian Reasoning and Machine Learning*. New York, NY, USA: Cambridge University Press

Bishop C, Tipping ME. 2003. Bayesian regression and classification. Advances in Learning Theory: Methods, Models and Applications

Blanton MR, Roweis S. 2007. *The Astronomical Journal* 133:734–754

Blei D. 2007. *COS 597C: Bayesian nonparametrics, Lecture 1*

Brammer GB, van Dokkum PG, Coppi P. 2008. *The Astrophysical Journal* 686:1503–1513

Del Pozzo W, Vecchio A, Berry CPL, Ghosh A, Haines TSF, Singer LP. 2018. *Monthly Notices of the Royal Astronomical Society* 479:601–614

Forrest B, Tran KVH, Broussard A, Cohn JH, Robert C. Kennicutt J, et al. 2018. *The Astrophysical Journal* 863:131

Garnett R. 2018. Active search in big data era

Garnett R, Ho S, Bird S, Schneider J. 2017. *Monthly Notices of the Royal Astronomical Society* 472:1850–1865

Garnett R, Ho S, Schneider J. 2015. *Finding Galaxies in the Shadows of Quasars with Gaussian Processes*. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15. JMLR.org

Gelman A, Carlin JB, Stern HS, Rubin DB. 2014. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd ed.

Kriek M, van Dokkum PG, Whitaker KE, Labbé I, Franx M, Brammer GB. 2011. *The Astrophysical Journal* 743:168

Pâris, I., Petitjean, P., Aubourg, É., Bailey, S., Ross, N. P., et al. 2012. *A&A* 548:A66

Rasmussen CE, Williams CKI. 2005. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press

Scoville N, Aussel H, Brusa M, Capak P, Carollo CM, et al. 2007. *The Astrophysical Journal Supplement Series* 172:1–8

Straatman CMS, Spitler LR, Quadri RF, Labbé I, Glazebrook K, et al. 2016. *The Astrophysical Journal* 830:51

Teh YW, Jordan MI, Beal MJ, Blei DM. 2006. *Journal of the American Statistical Association* 101:1566–1581

Whitaker KE, Labbé I, van Dokkum PG, Brammer G, Kriek M, et al. 2011. *The Astrophysical Journal* 735:86

Wolf, C., Meisenheimer, K., Rix, H.-W., Borch, A., Dye, S., Kleinheinrich, M. 2003. *A&A* 401:73–98