Draft version August 8, 2019
Preprint typeset using LaTeX style emulateapj v. 12/16/11

ON THE PROSPECTS OF ANALYZING STAR-FORMING GALAXY SPECTRA WITH MACHINE LEARNING MODELS

H. Jabran Zahid[1], I-Ting Ho[2], & Rolf-Peter Kudritzki[3]
[1]Smithsonian Astrophysical Observatory, Center for Astrophysics Harvard & Smithsonian - 60 Garden St., Cambridge, MA 02138, USA
[2]Galaxies and Cosmology Department - Max-Planck Institute for Astronomy - Königstuhl 17, 69117 Heidelberg, Germany and
[3]University of Hawaii at Manoa, Institute for Astronomy - 2680 Woodlawn Drive, Honolulu, HI 96822, USA
Draft version August 8, 2019

ABSTRACT

Subject headings: galaxies: evolution − galaxies: ISM − galaxies: formation − galaxies: abundances

1. INTRODUCTION

Spectroscopy is a fundamental tool of modern astrophysics and spectroscopic observations are the foundation of our modern understanding of galaxy formation and evolution. Galaxy spectra depend on the physical properties and conditions of galaxies and thus spectroscopic observations provide observational constraints for understanding the physics governing the formation and evolution of galaxies. However, galaxies are complex systems and thus extracting information about their physical properties from spectroscopic observations is not a trivial. Developing tools to analyze large astronomical data sets is a significant challenge which will remain pertinent given the size and complexity of the next generation of astrophysical data sets.

The challenge of spectroscopic analysis is to extract complex information about the physical properties and conditions of astrophysical systems which is encoded in a spectrum. One approach is to analyze specific regions of the spectrum which are known to be sensitive to particular properties. For example, star formation rates, metallicities and ISM physical conditions of galaxies are often derived from optical emission lines (Osterbrock 1989; Kennicutt & Evans 2012, e.g.,) when galaxies are actively forming stars. For quiescent galaxies, astronomers have used specific absorption lines to constrain age and metallicity (Worthey 1994). The advantage of such approaches is that they reduce the complexity of the data by narrowing the spectral range analyzed. However, the physical properties of interest are typically imprinted across the full spectral range and thus relevant information is discarded.

Recent developments in stellar population synthesis modeling (Conroy 2013) and computational power allow for more sophisticated approach of full spectral fitting (Cappellari & Emsellem 2004; Cid Fernandes et al. 2005). The full spectral fitting approach fits stellar population synthesis models for which stellar properties are known to observed spectra essentially decomposing a spectrum into a linear combination of individual stellar populations with varying properties. The advantage of such an approach is that the best-fit is straightforward to interpret because it is simply a linear combination of known quantities. The difficulty is that the parameters space is typically quite large and thus a minimization of many parameters can be computationally expensive.

Machine learning is an umbrella term for a broad class of algorithms which rely on the use of "training data" to build numerical models which are not explicitly defined. The modeling technique relies on pattern recognition to make inferences. Here we develop an multilayer perceptron artificial neural network (ANN) which is a machine learning algorithm inspired by biological neural networks found in animal brains. These ANNs are known to be universal function approximators (Cybenko 1989); they are able to represent a wide range of functions without explicit instructions, i.e. the machine learns the representation via training. Machine learning has been applied to analysis of astronomical data (Ball & Brunner 2010; Ivezić et al. 2014) though its application to extragalactic spectroscopy remains limited (e.g., Tao et al. 2018; Lovell et al. 2019; Ho 2019).

Our motivation for exploring machine learning as a spectroscopic analysis technique is that the physical properties of galaxies encode patterns in spectra. Given a set of spectra with known properties, we can train the machine to learn these patterns. Thus, a key ingredient for machine learning is the "training data" which is a set of inputs for which the output is known. We use stellar population synthesis to generate model spectra with explicitly defined star formation and chemical evolution histories as our training data. We explicitly define the input star formation and chemical evolution histories used to generate the spectrum and thus choose an output which can be derived from these inputs. Here we demonstrate the feasibility of this approach by training the ANN to output the luminosity weighted metallicity; more complex outputs are possible and should be explored in future studies.

Our aim is to demonstrate the feasibility of using artificial neural networks to analyze galaxy spectra. Here we build on the work of Zahid et al. (2017, Z17 hereafter). These authors derived the stellar mass - stellar metallicity of star-forming galaxies by fitting the stacked spectra from SDSS with a linear combination of single burst stellar population models of different ages and metallicities. Our approach here is to train the ANN using SPS models and then apply our trained model to the stacked SDSS data. We demonstrate that a ANN can recover an MZ relation consistent with a more standard χ^2- minimization approach used in Z17. Applying the χ^2- minimization approach of Z17 is computationally prohibitive. The advantage of the ANN is that once trained, the model can be trivially applied to a large number of spectra. We apply the model to derive stellar metallicities for 200,000 star forming galaxies in the SDSS.

We emphasize that this work is meant as a short, pilot study demonstrating the feasibility of using stellar population synthesis models to train a neural network to learn how to extract information from integrated galaxy spectra. Our aim is to motivate more research into the application of machine learning techniques to the analysis of spectroscopy.

2. SOFTWARE AND HARDWARE

We develop our model using Keras (Chollet et al. 2015), an open source neural-network library which serves as a high-level application programming interface (API) utilizing the open source TensorFlow (Abadi et al. 2015) library developed by Google for numerical computation used in deep learning. We implement these libraries in Python 3.6.5 which we run on a 2017 MacBook Pro with 3.1 GHz Intel Core i7 with 16 GB of Memory.

2.1. *Multi-layer perceptron*

We build a feed-forward artificial neural network (ANN) using the multi-layer perceptron (MLP) framework. A schematic of the network architecture is shown in Figure **??**. The MLP is constructed using Keras and TensorFlow. The fully-connected MLP has three components: input layer, hidden layer(s), and output layer.

We adopt a novel architecture specifically designed for interpretability. We split each pre-processed spectrum into 25 chunks of ~ 100 flux elements and feed each chunk into fully connected layer with 4 nodes. Each node of the fully connected layer is a perceptron. Each perceptron performs a simple calculation

$$p(x) = \sum_i w_i x_i + b, \tag{1}$$

where w_i and b (i.e. weight and bias) are the model variables and x_i are input from the previous layer. All layers use a sigmoid activation function that adds nonlinearity to the model except for the final layer which has a linear output. We tried the standard rectified linear units optimization but found that it did not generalize as well as the sigmoid output. The sigmoid activation function is:

$$A(x) = \frac{1}{1+e^{-x}}. \tag{2}$$

In other words, each neuron in the hidden layers ingests outputs of the previous layer, x_i, and performs the calculation

$$A(p(x)) = \frac{1}{1+e^{-\sum_i w_i x_i + b}}. \tag{3}$$

For the input layer, each node acts as the output of narrow pass-band filter where the filter response is learned by the network during training. Thus, the first layer of the network outputs 400 numbers, 4 for each of the 25 chunks which are then passed through two fully connected layers of x and y nodes respectively.

One challenge of ANNs is that information is reduced in complexity as it is passed through the network. The network essentially learns how to transform ~ 2500 numbers, i.e. flux elements into the output label, i.e. the luminosity weighted metallicity. Information passing through the network quickly becomes transformed such that the inputs/outputs of deep network layers are difficult to interpret. Our network architecture is designed such that the first input layer can be interpreted as analogous to the output of a narrow-band filter and the weights can be interpreted as the filter response function. The inputs/outputs of the deeper, hidden network layers are not straightforward to interpret.

3. OBSERVATIONAL DATA AND TRAINING DATA

3.1. *Observational Data*

We analyze the Main Galaxy Sample of the Sloan Digital Sky Survey (Abazajian et al. 2009; Alam et al. 2015). The parent sample consists of $\sim 900,000$ galaxies observed over 10,000 deg^2 down to a limiting magnitude of $r < 17.8$ at $0 \lesssim z \lesssim 0.3$. The observed spectral range is $3800 - 9200$Å with a nominal spectral resolution of $R \sim 1500$ at 5000Å (Smee et al. 2013). We adopt total stellar masses, total star formation rates and emission line fluxes measured by the MPA/JHU group[1] (Kauffmann et al. 2003a; Tremonti et al. 2004; Brinchmann et al. 2004; Salim et al. 2007).

We select galaxies following the criteria described in Andrews & Martini (2013). The sample is restricted to $0.027 < z < 0.25$ to ensure rest-frame wavelength coverage of the $[OII]\lambda3727$ and $[OII]\lambda7330$ emission lines. We remove galaxies classified as active galactic nuclei (AGN) using the Kauffmann et al. (2003b) classification based on the Baldwin et al. (1981) diagram. The classification is based on a ratio of emission line fluxes. Star-forming galaxies are classified as having line ratios

$$\log([OIII]\lambda5007/H\beta) > \frac{0.61}{\log([NII]\lambda6584/H\alpha) - 0.05} + 1.3 \tag{4}$$

and

$$\log([NII]\lambda6584/H\alpha) < -0.05. \tag{5}$$

We select galaxies with $H\beta$, $H\alpha$ and $[NII]\lambda6584$ emission lines observed with a signal-to-noise (S/N) ratio ≥ 5. These selection criteria yield a sample of $\sim 200,000$ star-forming galaxies.

The data are stacked in 0.1 dex bins of stellar mass with stellar masses ranging between $10^{8.5} - 10^{11} M_\odot$. The spectra are corrected for galactic extinction using the maps of Schlegel et al. (1998) and are shifted to rest-frame wavelengths using the measured redshift. They are linearly interpolated between $3700 - 7360$Å with a wavelength pixel resolution $\Delta\lambda = 1$Å and are normalized to the mean flux between $4400 - 4450$Å. The spectra are coadded by taking the mean flux of all spectra in the stellar mass bin at each wavelength pixel element. Each spectrum is equally weighted in the average. Systematics related to the stacking procedure are discussed in AM13.

We empirically determine the spectral resolution of each stacked spectrum from fitting strong emission lines and find that it increases with stellar mass. To avoid systematic effects caused by different spectral resolution, we convolve all spectra such that emission lines have the line width of the most massive galaxy stacked spectrum which is 330 km s^{-1} (full width at half maximum). S/N of individual spectra is often too low to apply this correction procedure to individual spectra. We apply the correction derived from stacked spectra to individual galaxy

[1] http://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/

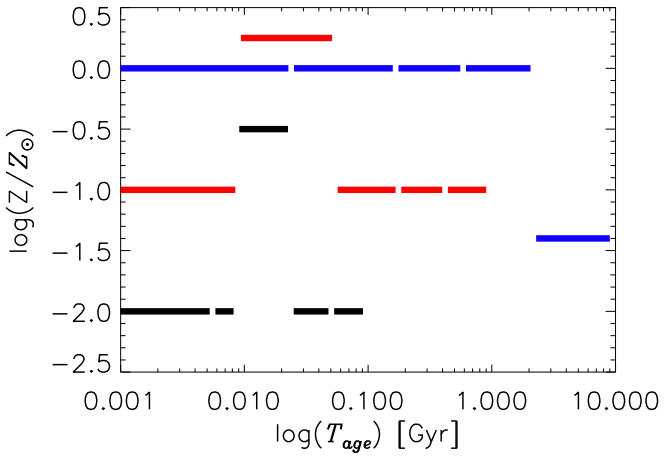

FIG. 1.—

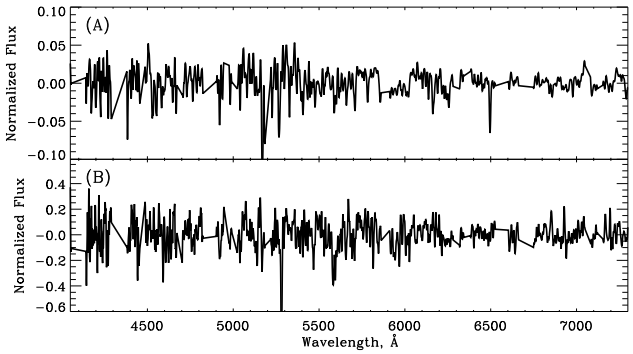

FIG. 2.—

spectra by interpolating the correction as a function of stellar mass.

3.2. *Training Data*

We calculate model spectra using the Flexible Stellar Population Synthesis (FSPS; v3.0) code (Conroy et al. 2009; Conroy & Gunn 2010). We adopt the Chabrier (2003) initial mass function, the Medium-resolution Isaac Newton Telescope Library of Empirical Spectra (MILES) library (Sánchez-Blázquez et al. 2006) and the Mesa Isochrones and Stellar Tracks (MIST; Dotter 2016; Choi et al. 2016). The intrinsic resolution of the model spectra is 2.5Å (Falcón-Barroso et al. 2011; Beifiori et al. 2011) which we convolve to match the resolution of the stacked spectra (i.e. 330 km s^{-1}).

We generate the training data using FSPS. We adopt a constant star formation history (SFH) varying the duration between 0.1-14 Gyr. The SFH duration grid consists of 21 values which are logarithmically spaced. Our metallicity grid consists of 21 metallicities evenly spaced between -2 to +0.4 $Z_{\odot}$ ($\Delta Z = 0.12$ dex). Our training data set consists of all permutations of the following prescription:

- Select SFH duration from grid of 21 values.
- Determine the five time intervals which contribute evenly to the luminosity of the final spectrum assuming constant SFR, Chabrier IMF and solar metallicity.
- Select metallicity from grid of 21 metallicities and assign to all five equal luminosity time intervals, thus producing a model of constant SFR and metallicity.
- Select one of five time intervals and assign a metallicity from grid, running through all possibilities.
- After completing above step, move to next of five equal luminosity time intervals and repeat above step.

Thus, we have 21 ages x 21 baseline metallicities x 21 single equal luminosity interval metallicity variations x 5 equal luminosity intervals yield a total of 46,305 training spectra. We label each model spectra by its luminosity weighted logarithmic metallicity and age.

Figure 1 shows examples of star formation and metallicity histories. These star formation and metallicity histories do not correspond to real galaxies, their primary purpose is to train the neural network with a spectra that systematically cover a range of ages and metallicities. Producing an appropriate training set is non-trivial and designing an appropriate procedure for generating training data requires additional research.

4. DATA PRE-PROCESSING

We process the observed and model spectra identically. We use the masking procedure described in Zahid et al. (2017). When fitting stacked spectra with models, we avoid the near ultraviolet part of the spectrum which is dominated by a series of absorption lines (Calcium H,K and Balmer series) and the [OII]λ3727 emission line and limit our analysis to the spectral range of 4050 − 7300Å. In this range we mask out a 30Å region centered on each ISM absorption and emission line. However, for the Balmer lines we mask out 75Å region to exclude the pressure broadened line profile wings. These masking windows are appropriate for the spectral resolution of the data and yield $\sim$ 2500 resolution elements.

We normalize stacked and model spectra to the pseudo-continuum by fitting a polynomial of degree 20. We divide the spectrum by the best-fit polynomial and subtract 1; thus the flux of each spectra is centered at zero. Figure 2 shows example pre-processed stacked and individual galaxy spectrum.

5. TRAINING PROCEDURE

The model spectra are noise free. We find that training on these data quickly leads to overfitting and the trained model generalizes poorly when applied to observed data. To mitigate overfitting and to improve generalization of the model, we add noise to a fraction of the training data. At each epoch we generate a noise spectrum by randomly drawing from a normal distribution of a fixed standard deviation. Thus, at each epoch we generate a new noise spectrum for each training spectrum which includes noise. Generating a new noise spectrum is necessary to prevent the network from learning noise features of any single spectrum thus mitigating overfitting.

We find that model generalization significantly improves when using a small amount L1 and L2 regular-

ization. We use L1 = X and L2 = Y. These values are adopted

Our observed spectra cover a range of S/N values. Thus, we split up our training data such that 20% of the input data is noise free and another 20% has a S/N of 5, 10, 20 and 50, respectively. This allows the model to train on data which cover a broad range in data quality. We pass 80% of the data through the network when training and reserve 20% for validation.

We train the model using the *Adam* (Kingma & Ba 2014) stochastic gradient-based algorithm and use the mean squared error as our cost or loss function. We found that the standard stochastic gradient descent (SGD) produces a more stable final result so we finish training with an additional 500 epochs using SGD. Thus, the model trains the x weights using gradient descent to minimize the mean squared difference between network prediction of metallicity and the luminosity weighted metallicity which is known for the model.

We find that model generalizes poorly if it is trained for too many epochs. Thus, we employ an early stopping criteria by assessing the results of the model applied to the stacked SDSS data. Essentially we compare the MZ relation recovered by our model to the MZ relation derived by Zahid et al. 2017 using a standard $\chi-^2$ minimization approach. This approach of early stopping is not ideal for several reasons. We assume that the MZ relation derived using χ^2 fitting procedure is robust and thus early stopping is essentially a "calibration" of our model results to that MZ relation. A more desirable procedure would converge to a solution thus avoiding the need for early stopping. A second related issue is that our stopping criteria is not objective, we essentially determine when to terminate training by eye.

Optimization of the model requires training many times whilst varying the hyperparameters and assessing by some objective criteria model scores. Training on a laptop takes x minutes. Thus, hyper parameter optimization requires high-performance computing and is beyond the scope of this work.

6. INTERPRETATION OF FIRST LAYER

7. APPLICATION TO SDSS AND THE MZR

Figure 3 shows the MZ relation derived from stacked data using our model (blue curve) compared to the one from Zahid et al. (2017) based on χ^2 fitting a linear combination single burst stellar population synthesis models. Within the systematic uncertainties, the two relations are consistent.

One major advantage of the machine learning model is the computational speed with which a model, once trained, can be used to infer metallicity of a spectrum. The $\chi-^2$ fitting approach takes several hours to converge for each spectrum. Thus, it is not feasible to apply this approach to determine metallicities of 200,000 individual galaxies. On the other hand our machine learning model can deliver metallicities of $\sim 200,000$ galaxies in less than one minute.

The blue curve in Figure 3 is the median MZ relation derived from the $\sim 200,000$ galaxies which were used to generate the stacked data. There is a clear offset between the stacked data MZ relation and the one derived from individual galaxies. We speculate that this is due to systematics related to the stacking procedure. Metallicty

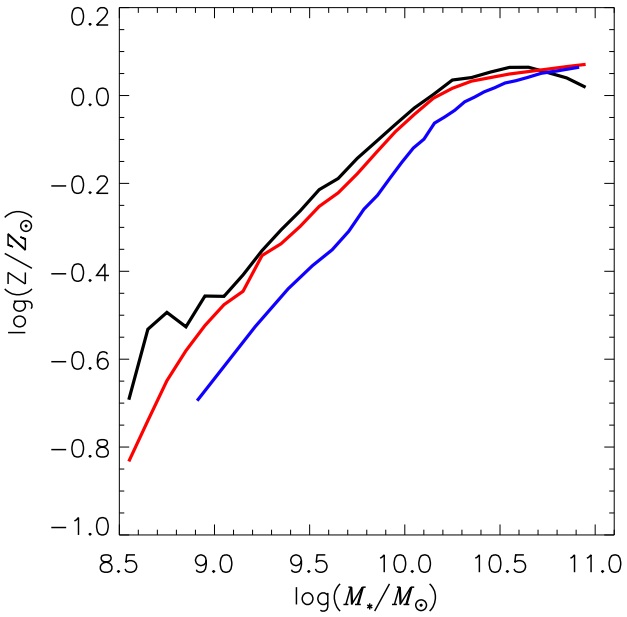

FIG. 3.— MZ relation determined by Zahid et al. (2017) by χ^2 fitting single burst stellar population synthesis models (black) curve compared with MZ relations inferred from the machine learning model developed in this study. The red curve is the MZ relation derived from stacked data and the blue curve is the median stellar metallicity $\sim 200,000$ galaxies calculated in equally populated bins of stellar mass.

information in the spectrum is encoded in the equivalent width of blended absorption lines. If metallicity the proportionality between metallicity and EWs is not log-linear, then the MZ relation derived from stacking spectra may be systematically different than the MZ relation derived from individual galaxies. On the other hand, the offset may also be due to unknown systematic effects. We tested and confirmed that resolution, flux calibration and varying S/N of data are not responsible for the offset but an exhaustive exploration of all systematics is beyond the scope of this work.

We test whether the stellar metallicities recover well known relation between stellar mass, star formation rate and and gas-phase oxygen abundance. At a fixed stellar mass, the gas-phase abundance is anti-correlated with star formation rate. At large stellar masses, the relation between star formation rate and metallicity at fixed stellar mass is reported to be either weaker, non-existent or positively correlated.

Figure 4 shows the relation between stellar mass, star formation rate and stellar metallicity. Similar to the gas-phase oxygen abundance, at a fixed stellar mass, the stellar metallicity is anti-correlated with star formation rate. We emphasize that we have trained our models on constant star formation histories and thus the recovery of this relation is a strong test that the model is "learning" and not jut "memorizing" how of infer metallicity from a galaxy spectra.

8. THE PROSPECTS OF MACHINE LEARNING FOR ANALYZING GALAXY SPECTRA

We discuss several directions which may be explored to more robustly develop machine learning modeling as a tool for statistical analysis of astronomical data. We

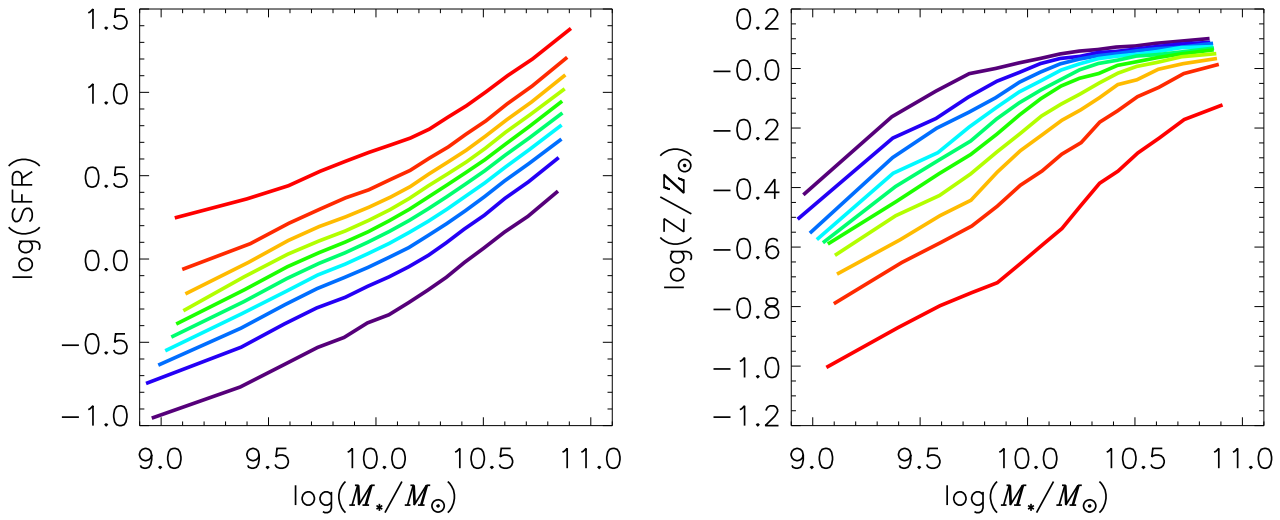

FIG. 4.— (A) Median star formation rate of $\sim 200,000$ galaxies first sorted in 15 equally populated bins of stellar mass and then sorted into 10 equally populated bins of SFR. (B) The median stellar metallicity corresponding to the sorted data in (A). The curves are color-coded the same, i.e., red and black curves indicate the high and low star formation rate galaxies in each stellar mass bin, respectively.

discuss some challenges we encountered and some of the alternative methods we explored that are promising but require more research to fully develop.

A fundamental challenge we faced was designing the training data set. We desire a training set that is representative of galaxies with a diverse set of star formation and chemical histories. The training data must be designed such that the alogrithm can "learn" how the star formation and metallicity variations impact the spectrum of a galaxy. However, we do not use star formation and metallicity histories that are empirically motivated (i.e. zahid et al. 2017 section blah) because we are concerned about circularity. The results of this study clearly demonstrate that machine learning models can "learn" how to extract metallciity from a galaxy spectrum that is consistent with more conventional techniques.

Here we focused on the luminosity weighted metallicity. We also use the luminosity weighted age of the population as a label, though analysis of galaxy ages is beyond the scope of this work. Part of the power of this modeling approach is flexibility. We are free to choose the desired outputs. Here we present on the luminosity-weighted metallicity but we tried additional tests in which we attempted to recover the star formation and metallicity history by using the metallicity and age of the five equal luminosity bins labels. Thus, we attempted to recover time resolved quantities related to the star formation and chemical evolution history. While we found some successes, we ultimately did not pursue this route due to limitations in our training data and label generation. A detailed study of what labels are most appropriate and the level of complexity that we can robustly recover from observations is warranted.

The particular model architecture we develop is designed such that we are able to interpret the first input layer as analogous to a narrow-band filter, thus providing a means to qualitatively assess the information content of galaxy spectra. Machine learning models could be developed to more rigorously quantify the information content of galaxy spectra. Such studies would be valuable for designing future observations, in particular large scale spectroscopic surveys.

We explored several additional machine learning models which could provide the basis for future efforts. In particular, we preprocessed the data using principle component analysis. We derived a basis set of spectra by performing a PCA on the single burst SPS models sampled at a range of ages and metallicities. We found that ~ 10 principle components accounted for nearly all the variance in the data. We then used these 10 components (i.e. orthonormal eigenvectors forming a basis set for all galaxy spectra) and projected each observed spectra on this basis set. Thus, each galaxy spectrum is represented by 10 numbers. These 10 numbers were then fed into various machine learning algorithms including a neural net, random forest and xgboost (Chen & Guestrin 2016). All these models were able to produce an MZ relation when applied to the stacked data. We choose to develop the model presented because it offered the greatest interpretability. An exhaustive exploration of various types of ML models is beyond the scope of this work but our experience suggests that these techniques should be explored.

Ensamble techniques in models are developed by combining many models have recently proven to be quite successful at tackling a range of modeling challenges. Given that we identify several classes of algorithms which are capable of recovering information from galaxy spectra, we expect that these ensamble machine learning techniques could be a valuable tool analysis of astronomical data. Further research in developing multiple models and combining them their outputs is warranted.

In this work, we analyze star-forming galaxies. Spectral analysis of quiescent galaxies is a significantly more mature area of study because of its simplicity. Incorporating quiescent galaxies into a single framework along with star forming galaxies would go a long way in unifying the analysis of these two populations. In particular,

multi-element abundances of quiescent galaxies are common. Such analyses applied to star-forming galaxies may be illuminating. A unified framework provides such an opportunity.

Developing realistic stellar population synthesis models remains a fundamental challenge of astrophysics. We assume that models generated using SPS are representative of observed galaxies. Given the nature of SPS techniques, this assumption breaks down at some level. SPS techniques have matured significantly in the previous decades but mismatch between models and observations remain. As the techniques for generating synthetic data using models improves, the use of these models for analyzing and interpreting observations becomes more robust.

9. CONCLUSIONS

Machine learning models are a well developed tool widely used in the information and technology industry. The power and flexibility of these techniques make them well suited for analyzing astronomical data sets. As computers become more powerful and astronomical data sets become larger, machine learning modeling will play a dominant role in astronomy. Our aim is to demonstrate a a succesful application of machine learning to a well studied problem. The success of our model demonstrates that machine learning models combined with SPS models are a powerful tool for analyzing galaxy spectra. The approach we develop is general enough that it may be applied to the study of stacked galaxy spectra of high S/N, the integrated spectra of individual galaxies, or to spatially resolved spectroscopy.

HJZ acknowledge the generous support of the Clay Fellowship.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, , , software available from tensorflow.org

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5

Ball, N. M., & Brunner, R. J. 2010, International Journal of Modern Physics D, 19, 1049

Beifiori, A., Maraston, C., Thomas, D., & Johansson, J. 2011, A&A, 531, A109

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151

Cappellari, M., & Emsellem, E. 2004, PASP, 116, 138

Chabrier, G. 2003, PASP, 115, 763

Chen, T., & Guestrin, C. 2016, in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, ACM, 785–794

Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102

Chollet, F., et al. 2015, Keras, https://keras.io, ,

Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, MNRAS, 358, 363

Conroy, C. 2013, ARA&A, 51, 393

Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Cybenko, G. 1989, Mathematics of control, signals and systems, 2, 303

Dotter, A. 2016, ApJS, 222, 8

Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95

Ho, I. T. 2019, MNRAS, 485, 3569

Ivezić, Ž., Connelly, A. J., Vand erPlas, J. T., & Gray, A. 2014, Statistics, Data Mining, and Machine Learning in Astronomy (Princeton University Press)

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003a, MNRAS, 341, 33

Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003b, MNRAS, 346, 1055

Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531

Kingma, D. P., & Ba, J. 2014, arXiv preprint arXiv:1412.6980

Lovell, C. C., Acquaviva, V., Thomas, P. A., et al. 2019, arXiv e-prints, arXiv:1903.10457

Osterbrock, D. E. 1989, Astrophysics of gaseous nebulae and active galactic nuclei (University Science Books)

Salim, S., Rich, R. M., Charlot, S., et al. 2007, ApJS, 173, 267

Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, AJ, 146, 32

Tao, Y., Zhang, Y., Cui, C., & Zhang, G. 2018, arXiv e-prints, arXiv:1801.04839

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Worthey, G. 1994, ApJS, 95, 107

Zahid, H. J., Kudritzki, R.-P., Conroy, C., Andrews, B., & Ho, I.-T. 2017, ApJ, 847, 18