The Astrophysical Journal, 901:152 (10pp), 2020 October 1 https://doi.org/10.3847/1538-4357/abb0e3

A Machine-learning Approach to Integral Field Unit Spectroscopy Observations. I. H II Region Kinematics

Carter Rhea[1], Laurie Rousseau-Nepton[2], Simon Prunet[2], Julie Hlavacek-Larrondo[1], and Sébastien Fabbro[3,4]
[1] Département de Physique, Université de Montréal, Succ. Centre-Ville, Montréal, Québec, H3C 3J7, Canada; carterrhea@astro.umontreal.ca
[2] Canada–France–Hawaii Telescope, Kamuela, HI, USA
[3] NRC Herzberg Astronomy and Astrophysics, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada
[4] Department of Physics and Astronomy, University of Victoria, Victoria, BC, V8P 5C2, Canada
Received 2020 June 8; revised 2020 August 17; accepted 2020 August 18; published 2020 October 5

Abstract

SITELLE is a novel integral field unit spectroscopy instrument that has an impressive spatial (11 by 11 arcmin), spectral coverage, and spectral resolution ($R \sim 1$–20,000). SIGNALS is anticipated to obtain deep observations (down to 3.6×10^{-17}erg s^{-1} cm^{-2}) of 40 galaxies, each needing complex and substantial time to extract spectral information. We present a method that uses Convolution Neural Networks (CNN) for estimating emission-line parameters in optical spectra obtained with SITELLE as part of the SIGNALS large program. Our algorithm is trained and tested on synthetic data representing typical emission spectra for H II regions based on Mexican Million Models database (3MdB) BOND simulations. The network's activation map demonstrates its ability to extract the dynamical (broadening and velocity) parameters from a set of five emission lines (e.g., Hα, N [II] doublet, and S [II] doublet) in the SN3 (651–685 nm) filter of SITELLE. Once trained, the algorithm was tested on real SITELLE observations in the SIGNALS program of one of the southwest fields of M33. The CNN recovers the dynamical parameters with an accuracy better than 5 km s^{-1} in regions with a signal-to-noise ratio greater than 15 over the Hα line. More importantly, our CNN method reduces calculation time by over an order of magnitude on the spectral cube with native spatial resolution when compared with standard fitting procedures. These results clearly illustrate the power of machine-learning algorithms for the use in future IFU-based missions. Subsequent work will explore the applicability of the methodology to other spectral parameters such as the flux of key emission lines.

Unified Astronomy Thesaurus concepts: H II regions (694); Emission nebulae (461); Convolutional neural networks (1938); Neural networks (1933); Supernova remnants (1667); Interstellar medium (847)

1. Introduction

H II regions lay the foundation of many studies from star formation in galaxies, to galactic evolution and cosmology, and are one of the main drivers of observational extragalactic astronomy (e.g., French 1980; Weedman et al. 1981; Veilleux & Osterbrock 1987). H II regions form when the gaseous clumps are irradiated by an interior young and hot star or cluster of stars causing the gas to become partially or completely ionized (e.g., Osterbrock & Ferland 1989; Shields 1990; Franco et al. 2000). They are primarily composed of hydrogen and helium, but contain nonnegligible amounts of metals and their ionized counterparts (e.g., Shields & Tinsley 1976; Kennicutt & Oey 1993; Oey & Kennicutt 1993; Garnett & Shields 1987). The characteristic bright emission lines coming from recombination and collision between the free electrons and the different atoms/ions in the nebulae are observed at large distances and allow the study of interstellar matter and its primary constituents (e.g., Baldwin et al. 1981; Crawford et al. 1999; Kewley et al. 2006). Additionally, the omnipresence of the H II regions in some galaxies allow for the study of galactic disk dynamics (e.g., Epinat et al. 2008), magnetic fields and turbulence at large- and small-scales (e.g., Odell 1986; Haverkorn et al. 2015; Beck et al. 1996; Quireza et al. 2006; Pavel & Clemens 2012), and the importance of various feedback mechanisms that inject energy into the ISM, i.e., stellar winds, supernovae, and radiation pressure (e.g., Ramachandran et al. 2018, 2019; McLeod et al. 2020).

More recently, the use of integral field spectroscopy on nearby galactic and extragalactic H II regions has offered a more complete view of their physical properties (e.g., Sánchez et al. 2012; Bundy et al. 2014; Leroy et al. 2016). Also, increasing spectral and spatial resolution has allowed for the study of the complex dynamical structures of the H II regions and pushed the limit of previous analysis methods meant for integrated/unresolved spectra of H II regions (e.g., Martins et al. 2010; Sánchez et al. 2012; Drissen et al. 2014). Typical fitting procedures used to extract the dynamics and emission-line flux measurements from H II region spectra require a good prior estimate of the velocity as well as the number of velocity components to be fitted (e.g., Sánchez et al. 2016; Bittner et al. 2019; Zeidler et al. 2019). Defining the range of those priors is usually not a problem when the ensemble of spectra shows similar characteristics. While the typical range of velocity seen in galactic disks can easily vary by a few hundred km s^{-1} (e.g., Dressler et al. 1983; Bregman 1980; Sancisi et al. 2008), and the internal dynamics of H II regions can add thermal/turbulent broadening and expansion velocity to the galactic contribution (e.g., Arsenault 1986; SOFUE 1995), the typical velocity prior for a given spectral data cube can be very broad and is often not precise enough to ensure a proper fit of the entire data set. We are additionally facing new challenges in the dynamical analysis, because the spatially resolved H II region spectra often contain emission from different phases of the ISM (along the line of sight) and can be composed of multiple dynamically distinct components (e.g., expanding shells; Relaño & Beckman 2005; Rozas et al. 2007), each having a different thermal/turbulent

broadening. Of course, fitting two or more components with the proper velocity and broadening priors is the best approach in such a case, but only when such components are actually present in the spectra (e.g., Le Coarer et al. 1993; Relaño et al. 2005).

Ultimately, extracting the information in a consistent manner from high spectral and spatial resolution data cubes requires a dedicated method to estimate the priors on the different spectral parameters, taking into account the variation of the observed spectral features across the field of view (FOV).

SITELLE, the Imaging Fourier Transform Spectrograph of the Canada–France–Hawaii Telescope, produces spectral data cubes containing over 4 million pixels with adjustable resolving power (up to 10,000) and has an instrumental line shape described by a sine cardinal function (Baril et al. 2016; Martin & Drissen 2017; Drissen et al. 2019). Its $11' \times 11'$ FOV contains more than 4 million pixels for which the spectral sampling and resolution varies as a function of their relative position angle with the mobile mirror. Moreover, emission-line intensities (and therefore line intensity ratios) may vary significantly across the parameter space of the physical properties observed in H II regions.

All together, these characteristics make a typical template fitting strategy (e.g., cross-correlation function maximization) very difficult to implement since the sine cardinal function side lobes affect neighboring line intensity and shape, and the position of the lobes with respect to the central position of the line varies with spectral resolution (changes across the FOV). In addition, the variation of line intensity ratios between different emission regions can lead to gross errors on the velocity estimates when a single template spectrum is used. Therefore, an adapted approach is developed here to solve these issues while still fitting entire data cubes, using the same uniform and reproducible method and including the dynamical and spectral complex nature of the resolved H II regions.

This paper explores the use of a Convolution Neural Network to resolve deficiencies in the existing fitting software ORCS—*Outils de Réduction de Cubes Spectraux*. Although the ORCS fitting routines are robust, they require a human-generated prior for all fits; this paper demonstrates the use of machine learning to estimate the priors with no human input. In Section 2, we outline the Convolution Neural Network and the synthetic data set used to train the network. We explore the success of our Convolutional Neural Network (CNN) to the synthetic data in Section 3. In Section 4, we discuss the applicability of our methodology to low resolution spectra. Additionally, we apply the CNN to a field of M33 in order to test its efficacy in real observations. Finally, in Section 5, we recap the main successes and outline our future work.

2. Methodology

2.1. Convolutional Neural Networks

Neural networks have been used extensively in astronomy to classify galaxies (Storrie-Lombardi et al. 1992), separate galaxies from stars (Bertin 1994), categorize dynamic parameters of galaxy clusters (e.g., Ntampaka et al. 2016, 2019), explore astrophysical morphologies at differing scales (e.g., Iwasaki et al. 2019; Sadaghiani et al. 2019), derive galaxy redshift from wide band images (Pasquet et al. 2019), and extract emission-line parameters from spectra (e.g., Baron 2019; Ucci et al. 2019; Olney et al. 2020). A recent effort to calculate the parameters of H II regions from their spectra, GAME,[5] employs a combination of Decision Trees and AdaBoost in order to predict physical parameters (Ucci et al. 2017, 2018). In lieu of this, our method uses a CNN architecture designed by Fabbro et al. (2018), monikered STARNET, which has already demonstrated success in estimating emission-line parameters from stellar spectra.

During the course of this work, we became aware of the work of Keown et al. (2019), which uses an approach similar to ours to estimate the velocity and broadening of high-resolution radio emission lines, taking into account possible multiple velocity components. While their work focuses on high-resolution, isolated emission lines, ours focuses on lower resolution spectra observed on a wide field of view, hence often with a wide velocity distribution. In addition, the SITELLE ILS extended structure prevents us in any case from considering the different emission lines separately.

Our convolutional neural network is graphically depicted in Figure 1 and laid out as follows:

1. 8×8 convolution with four filters
2. 4×4 convolution with eight filters
3. Global max pooling with four filters
4. 20% dropout
5. 256 fully connected nodes
6. 128 fully connected nodes
7. 2 output neurons

The CNN takes the normalized SITELLE emission spectra obtained with the SN3 filter (651–685 nm) and returns an estimate on the velocity ($\mathrm{km\ s^{-1}}$) of the lines and their broadening ($\mathrm{km\ s^{-1}}$), assuming they are consistent over the five major emission lines in SN3. We tested several scaling functions (RobustScaler, StandardScaler, and MinMaxScaler); although we obtained the tightest constraints with the MinMaxScaler, the activation map revealed fitting nonphysical features and noise. We therefore normalize the spectrum to have a maximum value equal to unity.

In order to ensure the appropriate hyper-parameters, we explored their spaces extensively using the random search algorithm, as implemented by sklearn, embedded in a tenfold cross-correlataion. Throughout our training, we saw no significant deviation from the results reported by Fabbro et al. (2018). Therefore, we adopted the same hyper-parameter values as used in the standard STARNET procedure. Structural hyper-parameters can be readily seen in Figure 1. In order to view the other parameters (i.e., learning rates, decay rates, etc.), we suggest the reader view our github page: https://github.com/sitelle-signals/Pomplemousse. We report a maximum number of 10 epochs and an initial batch size of eight spectra.

2.2. Synthetic Data

In order to demonstrate the feasibility of using a CNN to identify the correct spectral parameters, we construct a set of synthetic data on which to train and test the network. The synthetic data set used in this study was created using the ORB software developed to reduce data from SITELLE (Martin et al. 2016). To generate synthetic spectra, we use the ORB create_cm1_lines_model function, which requires a number of parameters that will be defined in this section. Since our tool was developed primarily for SITELLE's

[5] https://game.sns.it/

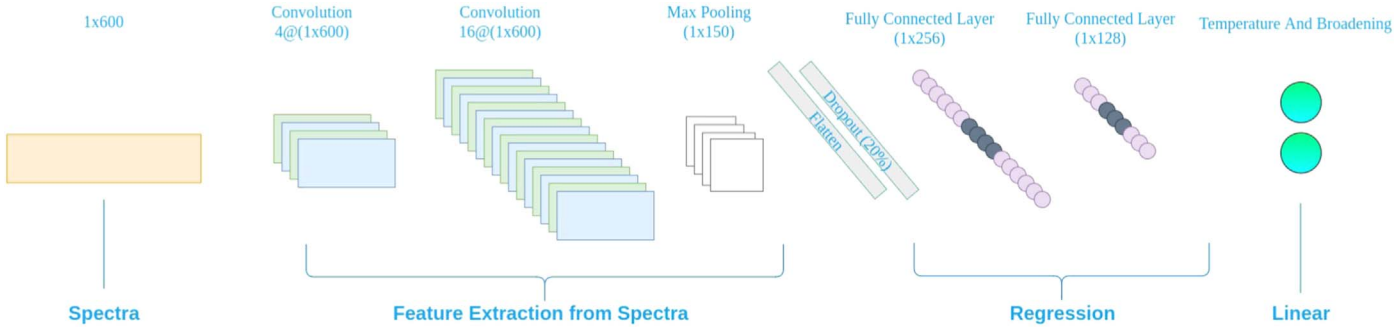

Figure 1. Cartoon of the convolutional neural network used in this work. As described in the text, it is an adaptation of the STARNET topology (Fabbro et al. 2018). The input spectra is first convolved in two separate layers before being condensed in a pooling layer. Once flattened, the vector is passed to two hidden layers. Finally, the velocity and broadening parameters are estimated using two separate output nodes denoted by the blue–green bar.

programs and the SIGNALS collaboration, we focused on the SN3-filter which covers a bandpass between 647 and 685 nm. In accordance with the SIGNALS survey, we select a primary spectral resolving power of 5000, an exposure time of 13.3 s per step, and 842 steps (Rousseau-Nepton et al. 2019). In order to replicate the change of spectral resolution across the cube, we allow the resolving power to randomly vary between 4800 and 5000 since the resolution will vary between these values in any given SN3 observation, which is a part of the SIGNALS program. We will model the following lines: [N II]λ6548, Hα (6563)Å, [N II]λ6583, [S II]λ6716, and S[II]λ6731. Furthermore, we use the `sincgauss` function as described in Martin et al. (2016) to include line broadening. We randomly varied the velocity between -200 and 200 km s^{-1}, while the broadening was randomly varied between 0 and 50 km s^{-1}. These ranges were selected from our prior knowledge of the distribution of velocities in M33 (Epinat et al. 2008) and the typical broadening in SITELLE data cubes at this spatial resolution. Note that we randomly selected the resolution, broadening, and velocity parameters with replacement for each synthetic spectrum. The final input required to construct the synthetic spectra is the amplitude of each emission line.

In order to calculate reasonable relative fluxes for the five lines while ensuring we are sampling the desired physical parameter space, we used the 3MdB[6]—Mexican Million Models Database (Morisset et al. 2015). The 3Mdb contains models created using the `CLOUDY v17.01` photoionization code based on a preselected set of emission region parameters and underlying ioinizing stellar spectra (Ferland et al. 2017). We use the `BOND` data set described in Vale et al. 2016, which contains spectra from H II regions similar to those expected to be found in SIGNALS. The `BOND` data set contains 63,000 spectra. Though the data set covers the physical parameter space of the emission nebulae we wish to study, it also contains a number of models that are outside the scope of our study. We describe varying parameters used in Table 1. While the `BOND` simulations have two simulation geometries, completely filled and thin shell, we remove all thin shell (fraction = 0.03) simulations from our sample. This leaves us with filled spheres with a density of approximately 100 cm^3 and represents a younger population of H II regions (e.g., Cedrés et al. 2013; Stasińska et al. 2015; Vale et al. 2016).

We further constrained the ionization parameter, U, and metallicity proxy, 12+log(O/H), to focus on SIGNALS-type H II regions (e.g., Kashino & Inoue 2019; Pérez-Montero et al. 2019; Rousseau-Nepton et al. 2019; Zinchenko et al. 2019). With these constraints, we extracted the amplitudes of the five emission lines present in SN3, first randomly selecting a model that passed our selection criteria. We then normalized the amplitudes with respect to Hα. After combining the five lines (with the appropriate instrumental line shape) and the simulated continuum emission, we add a noise component. The S/N is sampled from a uniform distribution between 5 and 30. Below an S/N of 5, the lines are nearly indistinguishable and the side lobes of the ILS are completely obstructed. We expect a nominal high (>20) S/N for Hα in the SIGNALS program. S/N effects will be investigated later in the article. Figure 2 shows a sample spectrum. At this stage, we create 50,000 mock spectra in the form of `FITS` files, which

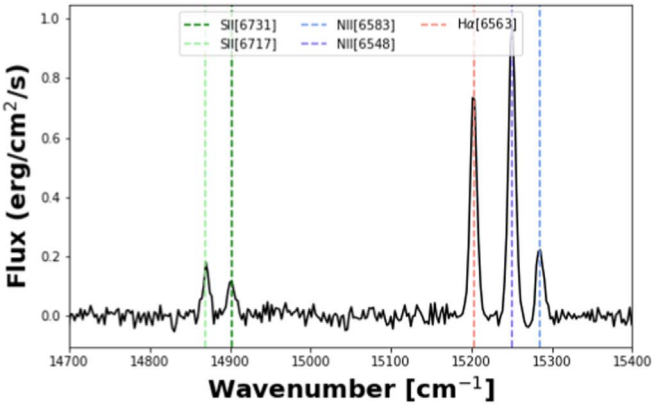

Figure 2. Example spectrum simulated using the process described in Section 2.2. As our population statistics suggest, this is not the only expected spectral shape. However, it is representative of the sample and clearly demonstrates the five emission-line peaks. This is the SN3 spectral coverage of SITELLE.

Table 1
H II Region Parameter Selection Used During the M3db Runs of the `BOND` Simulations

Parameter	Lower Limit	Upper Limit	Step Size
log(U)	−3.5	−2.5	0.5
Age (Myr)	1	6	1
12 + log(O/H)	7.4	9.0	0.2
log(N/O)	−2	0	0.5

Note. The initial run parameters were cut further in order to focus on the emission expected in the SIGNALS program. The step sizes were set by the 3Mdb runs (see Morisset et al. 2015 for more information).

[6] https://sites.google.com/site/mexicanmillionmodels/

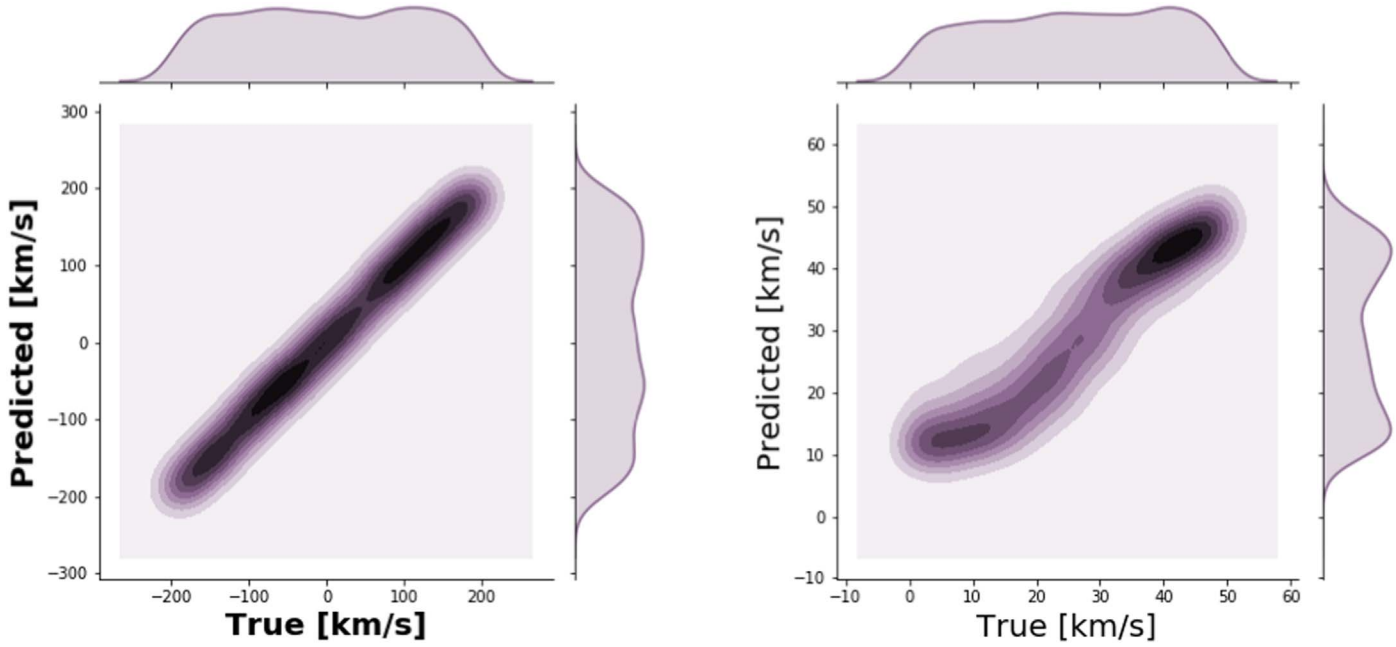

Figure 3. Kernel density estimation (KDE) plots for the test set. Left: true vs. predicted velocity values in km s^{-1}. Right: true vs. predicted broadening values in km s^{-1}. In both plots we can see that the predicted values accurately mimic the true values. Note the change in scales between the two plots.

contain the emission parameter information (e.g., velocity, broadening, and resolution).

2.3. SITELLE Data

2.3.1. Calibration and Data Reduction

Observations of M33 were taken during the Queued Service Observing period 18B (Program 18BP41, P.I. Laurie Rousseau-Nepton) at the Canada–France–Hawaii Telescope on the summit of Maunakea, Hawaii, using SITELLE. These exposures were taken with the SN3 filter, which covers a range from 651 to 685 nm for a total of 4 hr with a spectral resolving power of $R \sim 5000$. The pointing was centered on a single field in M33 and is part of a larger observation of M33 in its entirety. This observation also forms a basis for the SIGNALS program, lead by Laurie Rousseau-Nepton, which aims to further categorize H II and star-forming regions in nearby galaxies. We note that the authors of this paper are members of the SIGNALS collaboration.

The raw data were reduced and calibrated using SITELLE's personalized software, ORBS (version 3.1.2 Martin et al. 2016). We are able to resolve five spectral emission lines from our observations: [S II]λ6713, [S II]λ6731, [N II]λ6548, Hα, [N II]λ6584. Using the function `SpectralCube.Map_Sky_Velocity()`, we fit the OH sky line velocities, assumed at rest w.r.t. the observer, with a geometric model of the interferometer; afterwards, we used the function `SpectralCube.Correct_Wavelength()` to refine the wavelength calibration of our data cube using the OH-line fit.

3. Results

In this section we apply our convolutional neural network outlined in Section 2.1 to our synthetic spectra with a resolution R∼5000. We retained 70% (35,000) of the spectra as our training set, 20% (10,000) as our validation set, and the remaining 10% (5000) as the test set (e.g., Tetko & Villa 1997). Training and validating our algorithm results in over 95% accuracy for both predicted parameters: the velocity and the broadening. Accuracy is defined as the ratio of correct parameter estimations to the total number of estimates. An estimate is considered correct if it agrees with the ground truth value up to two digits after the decimal (i.e., to the hundredth place). The combined mean absolute error, another common metric for regression tasks, is 5 km s^{-1}. Figures 3, 4, and 5 visually depict the accuracy of the CNN on the test set and the associated residuals, respectively. As the figures depict, the algorithm was well trained and is able to accurately predict both the velocity and the spectral broadening. As evidenced in Figures 3 and 4, the predicted values are close to the ground truth values. The KDE plots in Figure 3 demonstrate that the parameter space is being well sampled for both the velocity and broadening. Figure 5 demonstrates the Gaussian distribution of errors about zero; although the right panel reveals the slightly skewed error distribution of the broadening parameter, the shape is globally Gaussian and any distortion is believed to be caused by asymmetries within the training set. We report a standard deviation of ∼5 km s^{-1} for the velocity parameter.

This is well within the required limits as described in Martin et al. (2016) and Rousseau-Nepton et al. (2019) for an initial guess to be supplied to the `ORCS` software. The velocity error is required to be less than the channel width which corresponds to approximately 40 km s^{-1} for a resolution of 5000. The standard deviation of the broadening parameter is ∼5.5 km s^{-1}. Since SITELLE resolves the broadening parameter down to approximately 3 km s^{-1} for high S/N regions (∼1000), our broadening errors are near SITELLE's resolving power.

In order to compare the network results to those recovered by the `ORB/ORCS` software, we fit the test set using the `fit_lines_in_spectrum` routine. The velocity and broadening parameters were initialized as the precise velocity and broadening parameters used to construct the spectra. Although this is improbable to occur during a standard fitting procedure, hence the need for an accurate estimate, this demonstrates the

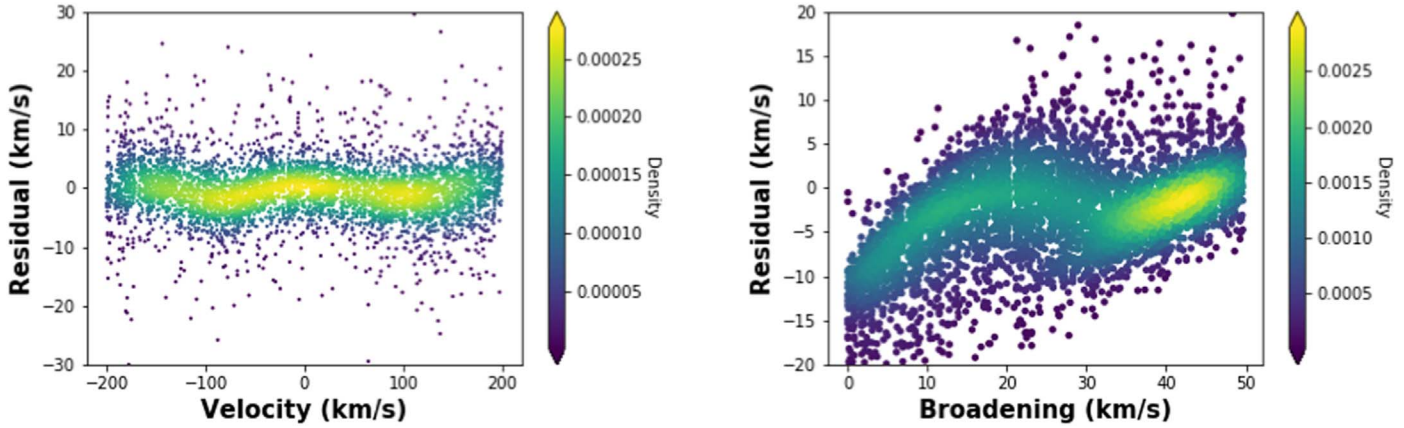

Figure 4. Left: velocity residual as a function of the true velocity. Although there exists a background substructure, it only affects a fraction of a percent of the total test set and is thus negligible. Right: broadening residual as a function of the true broadening. The pattern demonstrates a bias for low broadening values that is likely caused by the network's inability to distinguish a low amount of broadening. Moreover, the broadening naturally segregates itself into two physical peaks typical of H II regions and supernovae remnants, respectively (e.g., Veilleux & Osterbrock 1987; Vasiliev et al. 2015).

best possible case for the fitting algorithm. All other parameters were also set to those used to simulate the spectra. The fitting procedure recovers the true velocity with a standard deviation of $\sim$3 km s^{-1} and the broadening with a standard deviation of $\sim$4 km s^{-1}. Comparing these standard deviations with those from the CNN, we note that the `ORB/ORCS` recover the true parameters with marginally better accuracy.

Although the spread of errors shown in the Figures 5 and 4 do not reveal overt overfitting, we applied a standard *k*-fold cross-validation algorithm on 10 partitions of the training, validation, and test data (e.g., Picard & Cook 1984; Bengio & Grandvalet 2004). Overfitting occurs when the neural network learns the training set too well and is unable to generalize to other data sets such as the test set. Overfitting would manifest itself in these figures if they demonstrated a large spread of residuals (i.e., large errors). We also implemented a modified *k*-fold cross-validation algorithm in which we varied only the training and validation data while retaining the same test set. We report approximately the same accuracy values (within 5%) regardless of the fold and cross-validation technique. This further indicates the absence of overfitting (e.g., Molinaro et al. 2005; Cawley & Talbot 2010).

Additionally, we created an saliency map of our example spectrum from Figure 2, which can be seen with the filled circles in Figure 6. The saliency map delineates the regions of the input (in this case the spectrum) used by the convolutional neural network to learn (e.g., Simonyan et al. 2013) by calculating the gradient of the output with respect to the input. More precisely, the map is created by varying one input variable at a time and calculating the change in the loss function. In this manner the algorithm highlights the most important input nodes. We can clearly see by the clustering of data points in the image around the Hα and [N II]λ6548 lines that the network considers these lines to be the most important components for determining the velocity and broadening. This is consistent with our expectations since these two lines, unlike the others, are consistently above the continuum in H II regions. It is sensible that the network does not weigh the [S II] doublet heavily since they are often unobservable due to noise. Moreover, the network does not focus only on the peaks of the Hα and [N II]λ6548 lines, but also on their base. This indicates that the widening of the lines—which is directly affected by the velocity and broadening components—plays a crucial role in parameter estimation, as expected.

4. Discussion

While in Section 3, we demonstrated that the CNN algorithm is capable of extracting the correct spectral parameters (velocity and broadening) of the Hα, N [II], and S [II] lines for synthetic SITELLE observations, in this section, we examine the versatility of the model and its robustness when applied to real SITELLE observations. We also discuss the novelty of using such CNN algorithms for IFU observations in general (i.e., from other telescopes, especially in the context of upcoming 30 and 40 m class telescopes.

4.1. Versatility of the Model

While this technique is developed for the SIGNALS collaboration science case, aiming to obtain IFU observations of dozens of nearby galaxies, and thus $R \sim 5000$, we demonstrate its applicability to other studies of H II regions using SITELLE at various spectral resolutions. Since there exists a number of other SN3 observations, which are not a part of the SIGNALS program, that were taken with an average spectral resolving power near $R \sim 2000$, we wished to directly test our existing network and weights against synthetic data created with $R \sim 2000$ (e.g., Gendron-Marsolais et al. 2018; Rousseau-Nepton et al. 2018; Puertas et al. 2019). However, since the resolution sets the number of steps (i.e., data points) in our spectrum, a reduction of the resolution affects the length of the input data. In order to feed lower resolution spectra into our CNN, we would be required to smooth or interpolate the data so that we would have an input of an equivalent length—a requisite for use in a CNN. In doing so, we would be assuming a form of the interpolation (i.e., linear, a higher-order polynomial, spline, etc.), which might inject nonphysical and potentially biased information into the spectra (Horowitz 1974; Scargle 1982; Schulz & Stattegger 1997). We therefore do not modify the spectra, but instead we create an entirely new set of training, validation, and test data using the same routines employed to create our high spectral resolution synthetic data set with a resolution set to $R \sim 2000$.

After creating 30,000 synthetic spectra with a lower spectral-resolution, we divided the set into the training (70%),

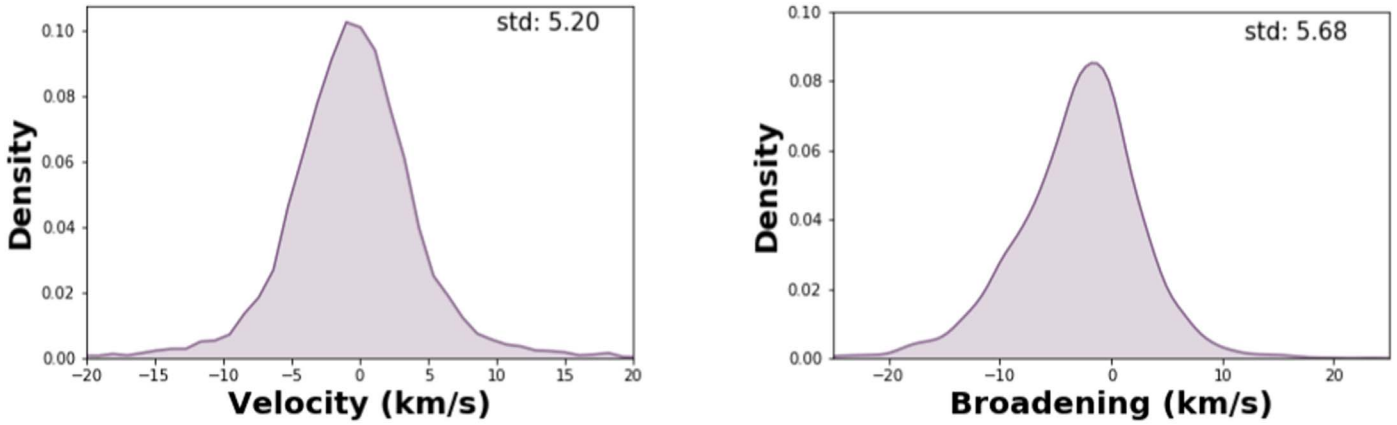

Figure 5. Left: density plot of the velocity residuals in km s^{-1} along with the standard deviation. Right: density plot of the broadening residuals in km s^{-1} in addition to the standard deviation. The asymmetry is likely due to the diversity of resolving power introduced in the training set.

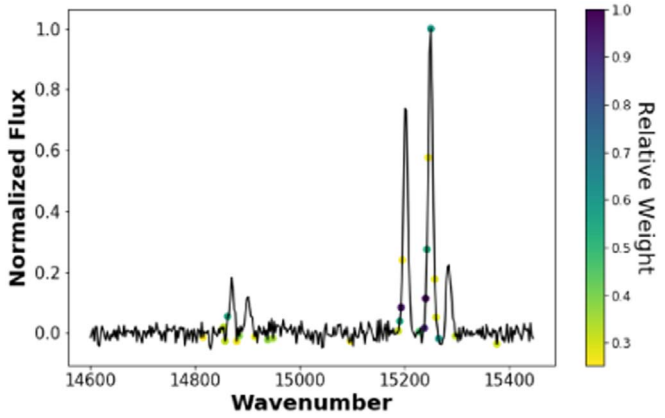

Figure 6. Activation or saliency map of our convolutional neural network applied to an example spectrum. The colored points represent the exact locations of the nodes in the input spectrum. Their color indicates their relative weight in the network. Weights under 0.25 are not shown for clarity.

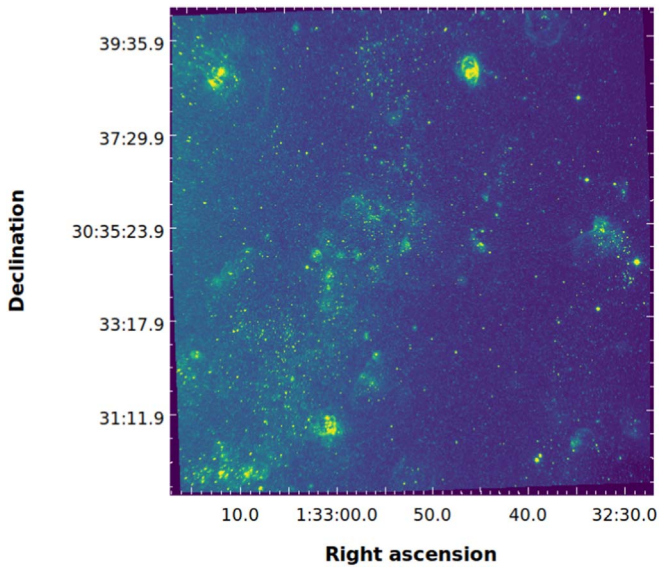

Figure 7. Deep, coadded SITELLE observation (4 hr) of M33 Field 7 using the SN3 filter. The image illustrates the density of emission-line regions in the outskirts of M33.

validation (20%), and test (10%) sets. After training and validating our convolutional neural network, we applied it to our test data. We report a nominal accuracy of both predictors (velocity and broadening) of 92% compared to 95% in the case of $R \sim 5000$. The standard deviation of the errors for the velocity and broadening are 75 and 12 km s^{-1}, respectively. We ran both k-fold cross-validation algorithms and again found consistency across the accuracy predictors. The results are coherent with our supposition that the method would extend well to relatively low resolution spectra since, even at $R \sim 2000$, we are able to reasonably resolve the emission lines. The reduced accuracy is reasonable since the emission lines are less well-resolved.

We attempted to use the network to predict low resolution SITELLE spectra ($R \sim 1000$); however, at this resolution, the lines are often indistinguishable and the algorithm fails to achieve high-fidelity results. Typical SITELLE's observing strategy for targets in the local universe and for the SIGNALS project, have an increased spectral resolution for the Hα filter (SN3) and often a lower resolution for other filters (typically $R \sim 1000$). The dynamical priors (velocity and broadening) can then be estimated using the higher resolution SN3 filter and applied on the other observations of the same field with the other filters. Overall, our results demonstrate that a CNN network is capable of reliably estimating spectral parameters (velocity and broadening) in SITELLE synthetic observations at high ($R = 5000$) and low ($R = 2000$) resolution, but that beyond $R = 1000$–1500, it fails because of the poor quality of observations. In other words, these results not only demonstrate that machine-learning algorithms can be used to estimate kinematic parameters, but they also demonstrate the technique's limitations.

4.2. Validation on a Real Data Set: The Case of M33

With the ability of the CNN to predict velocity and broadening parameters accurately for synthetic data, we apply our methodology to an emission region of M33's southeast field (Figure 7). This region is an excellent test-bed for our algorithm since it contains several types of emission regions (i.e., H II region, planetary nebulae, etc.) and is part of the SIGNALS survey.

Fits were calculated using the `ORCS fit_lines_in_region()` command centered on our five lines. Each grouping ([S II]λ6713/[S II]λ6731, [N II]λ6548/[N II]λ6584, and Hα) was fit simultaneously with a Gaussian convolved with a sinc function following the standard SITELLE procedure (Martin &

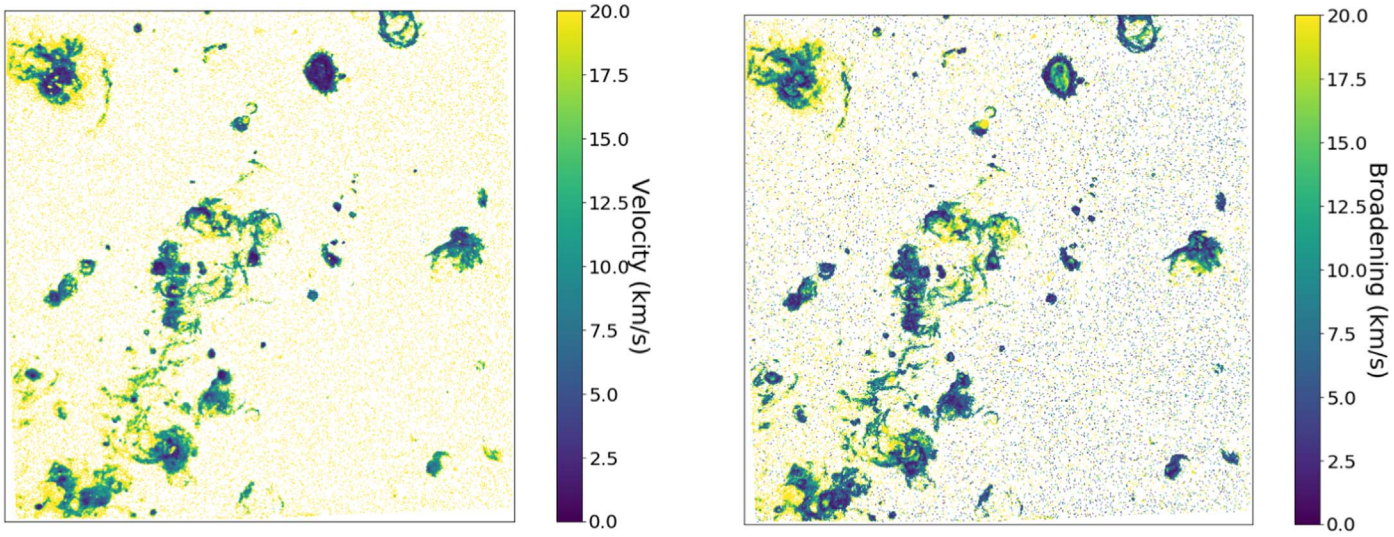

Figure 8. Left: residual map of the velocity calculated from the absolute difference between the final ORCS fit and the machine-learning priors calculated on an unbinned cube. Right: residual map of the broadening calculated from the absolute difference between the final ORCS fit and the machine-learning priors calculated on an unbinned cube. Both maps were smoothed using a two-dimensional Gaussian kernel with a sigma value equal to 2 pixels.

Drissen 2017). All lines were tied together with respect to the velocity and broadening. Fits were optimized using the Levenberg–Marquardt least-squares minimization algorithm. In order to execute a fit in ORCS, the user is required to input an initial guess for the velocity and broadening parameters; this is due to the nature of the minimization algorithm. The first set of priors was created by initially binning our cube into spatial bins of 8×8 followed by the standard ORCS fitting procedure. This standard method still requires an initial guess that the user must input. However, the machine-learning method for determining priors does not require any user input and can be applied directly on the unbinned data. All fits were run using a computing server located at the CFHT headquarters in Waimea, Hawaii, named *iolani*. The server has 2 Intel XEON E5-2630 v3 CPUs operating at 2.40 GHz with eight cores each. The configuration also has 64 GB of RAM available for computing purposes.

A key benefit of the machine-learning prior fits over the standard procedure is the economy of time associated with the machine-learning algorithm. Since no fitting and iterating is necessary, the calculation time scales approximately linearly with the number of spectra. Using a coarse initial binning, 8×8, the standard algorithm to calculate the priors takes approximately 4 hr in order to cover the entire cube. However, the unparallelized machine-learning algorithm takes only 180 s[7] to cover the same binned cube. Hence the machine-learning algorithm calculates the priors more than 100 times faster than the standard algorithm. We also calculate the time the machine-learning algorithm takes to estimate the velocity and broadening parameters for an unbinned cube; this takes approximately 4 hr—the same amount of time to calculate the standard priors on an binned (8×8) cube.

In addition to being considerably faster when estimating the priors, the machine-learning algorithm also obtains accurate estimates. In order to quantify this notion, we calculate the residual values over the cube between the unbinned final fits—using an 8×8 machine-learning prior—and the unbinned machine-learning estimates. We only retained pixels for the residual analysis, which demonstrated a flux value above our threshold of 2×10^{-17} erg s^{-1}. This threshold was chosen since it masks out all nan values and maintains the regions with clear emission. Figure 8 demonstrates that the residuals are low in central parts of the emission regions, where the signal-to-noise is high, while the residuals are higher in the outskirts where the signal-to-noise is low. This is likely due to the fact that our synthetic data was created using a high signal-to-noise ratio of 50; we will explore the effects of the S/N ratio in a future paper. While it is often desirable to study the emission in the outskirts in addition to the central emission, the low-residual regions outline locations of high-fidelity fits. In order to recover the velocity and broadening parameters in these regions, the machine-learning estimates on either the binned or unbinned cube can be used as priors for a standard ORCS fit. Moreover, since the standard prior calculation requires binning spatially, substructure information is inherently lost in these priors. On the other hand, the convolutional neural network priors do not require any binning and thus retain all structural spatial information.

Although we do not study all the complexities of the S/N impact on our CNN in this article, we include a short discussion on it here. We calculate the S/N by dividing the Hα flux by its fit uncertainty as calculated in our final ORCS fit. Although this is not exactly the S/N, it acts as a proxy value. With the residual maps and the S/N proxy map, we have the residual and signal-to-noise information for each pixel. We then binned residuals by signal-to-noise ratio with a step size of 1 between 5 and 20. Twenty is the maximum value of the S/N proxy and below 5 we do not see any coherent structure in the spectra. We culled outliers that were outside of the 3σ range. Finally, we calculated the median absolute residual and standard deviation in each S/N bin. As evidenced by Figure 9, the accuracy of the CNN increases as the signal-to-noise ratio rises, an expected trend. Figure 10 demonstrates that the broadening residual plateaus at an S/N of approximately 12; moreover, the figure indicates a discordance between the CNN's estimations and those obtained from ORCS fits. We believe this behavior is due to the presence of multiple emission components serendipitously located in high S/N regions

[7] Assuming a near-perfect speedup, we expect the parallelized algorithm to take approximately 25 s to run on *iolani*.

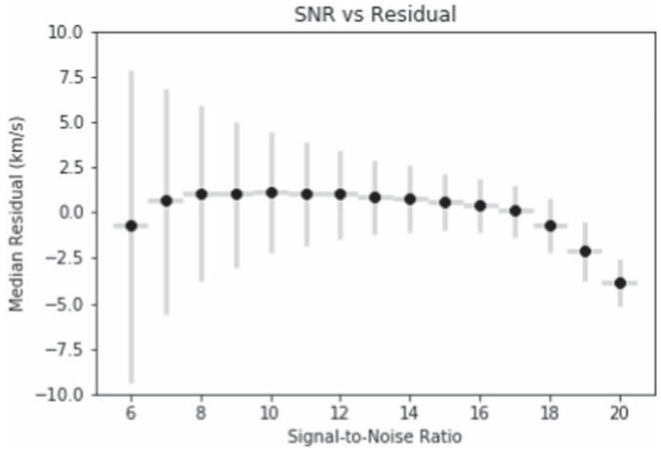

Figure 9. Proxy signal-to-noise ratio vs. mean absolute velocity residual (km s^{-1}) for the southwest field of M33. For each S/N bin, we excluded outliers before calculating the mean absolute residual and standard deviation (gray *y*-axis error bars). Each S/N bin has a width of 1.

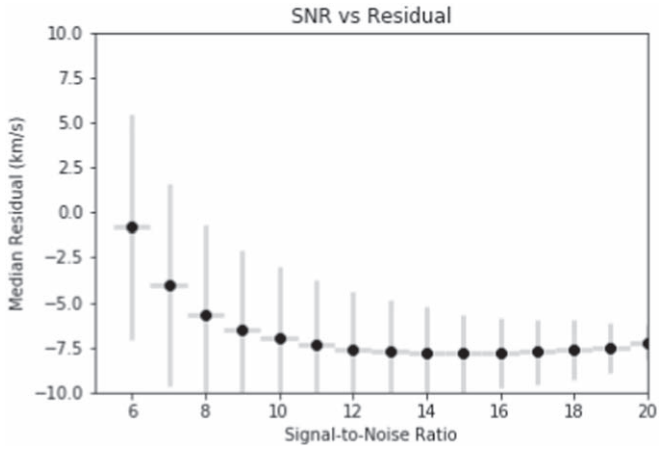

Figure 10. Proxy signal-to-noise ratio vs. mean absolute broadening residual (km s^{-1}) for the southwest field of M33. For each S/N bin, we excluded outliers before calculating the mean absolute residual and standard deviation (gray *y*-axis error bars). Each S/N bin has a width of 1.

(see the Appendix for a discussion). Multiple components affect the broadening parameter stronger than the velocity estimates. Even in standard fitting procedures, this poses a serious issue.

4.3. Universal Applicability

The methodology described in this paper is not limited to SITELLE data cubes. Indeed, the methodology naturally lends itself to any IFU-like data cube in which the observer has access to high-resolution spectral data such as the *K*-band Multi Object Spectrograph, KMOS (e.g., Sharples et al. 2013), or the Multi Unit Spectroscopic Explorer, MUSE (e.g., Bacon et al. 2010). Since the machine-learning algorithm is able to achieve reasonable estimations of the kinetic parameters (velocity and broadening) in a fraction of the time the standard fitting procedures take, it will play a crucial role in upcoming missions aimed at completing large-scale surveys using IFUs such as the Near-Infrared Spectrograph, NIRSpec (e.g., de Oliveira et al. 2018), on the James Webb Space Telescope and the MEGARA—Multi-Espectrógrafo en GTC de Alta Resolución para Astronomía—instrument on the Gran Telescopio Canarias (e.g., Gil de Paz et al. 2012).

5. Conclusions

A convolution neural network has been exploited in several astronomical applications ranging from dynamic mass estimates of galaxy clusters (e.g., Ntampaka et al. 2019) to the extraction of spectral parameters (e.g., Fabbro et al. 2018). This work applies a modified `STARNET` architecture (Fabbro et al. 2018) to high-resolution ($R < 2000$) SITELLE observations of H II regions in order to estimate the velocity and broadening parameters. Training, validation, and testing the machine-learning algorithm with synthetic data integrating the 3Mdb database (Morisset et al. 2015) demonstrates the feasibility of the method. We demonstrate that the algorithm fails to predict the spectral parameters for low resolution ($R \simeq 1000$) observations. We believe this is due to the lack of resolved spectral information resulting in partial blending of the main emission lines. However, above $R \sim 2000$, we are able to disentangle the lines better. We apply the convolutional neural network to the southwest field of M33 to calculate the velocity and broadening priors. Compared to the standard method for computing the priors, our method is over 100 times faster. Additionally, the machine-learning algorithm can reliably estimate the emission-line parameters for the entire unbinned cube in roughly the same amount of time it takes the standard algorithm to calculate the priors on an 8×8 binned cube.

The work presented here represents the first in a series of articles on the applications of machine learning to SITELLE spectra. In a subsequent article, we will present our work on the effects of the signal-to-noise ratio on convolution neural networks and how to mitigate the negative impacts.

We will also demonstrate the applicability of our methodology to calculate the fluxes (and ratios thereof) of emission lines, which will allow for the rapid classification of emission regions through grids of photoionization models (e.g., 3MdB). In the third proposed paper of the series, we will describe a machine-learning methodology to identify possible multiple, blended components within emission lines.

The authors would like to thank the Canada–France–Hawaii Telescope (CFHT), which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. The observations at the CFHT were performed with care and respect from the summit of Maunakea, which is a significant cultural and historic site. C.R. acknowledges financial support from the physics department of the Université de Montréal. J.H.-L. acknowledges support from NSERC via the Discovery grant program, as well as the Canada Research Chair program.

The progamming aspects of this paper were completed thanks to the following packages of the python programming language (Van Rossum & Drake 2009): `numpy` (van der Walt et al. 2011, `scipy` (Virtanen et al. 2020), `matplotlib` (Hunter 2007), `pandas` (McKinney 2010), `seaborn` (Waskom et al. 2017), `astropy` (Robitaille et al. 2013; Price-Whelan et al. 2018), `tensorflow` (Abadi et al. 2016, and `keras` (Chollet 2015).

Appendix
S/N and the Residual

As noted in Section 4.2, the broadening parameter (and the velocity parameter to a much lesser extent) exhibits an unexpected trend in its S/N versus residual plot (Figure 10). In this section, we explore potential reasons for this behavior: a

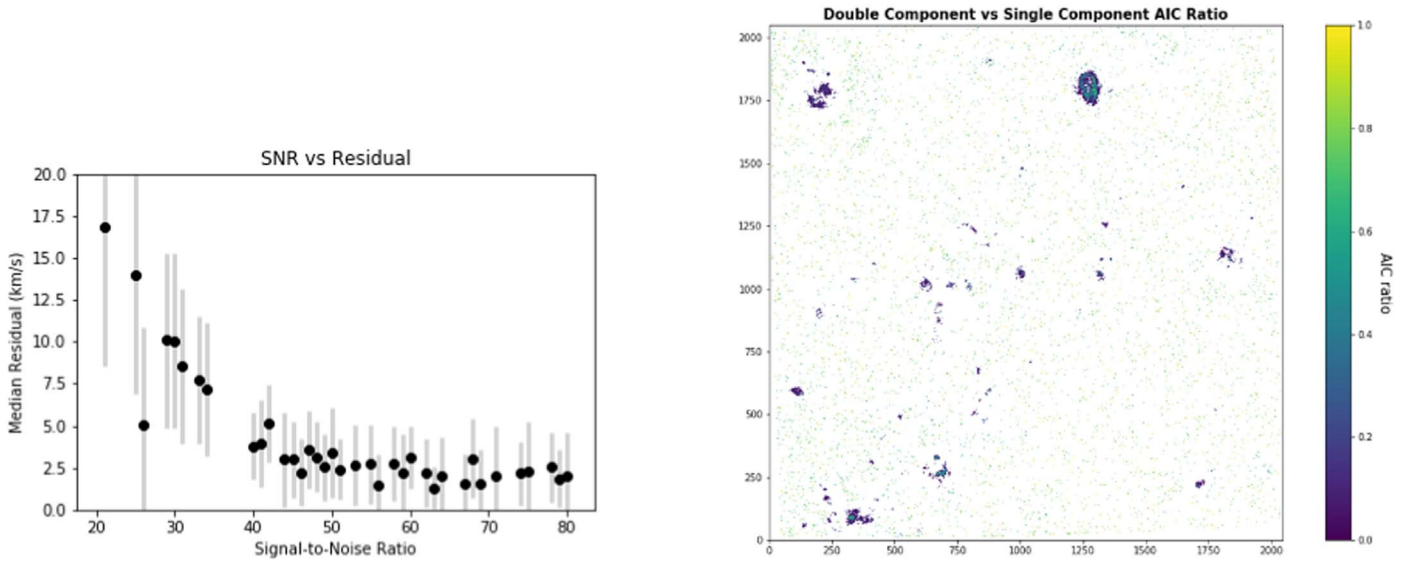

Figure A1. Left: proxy signal-to-noise ratio vs. mean absolute broadening residual (km s^{-1}) for synthetic data created to simulate a range of S/N values. For each S/N bin, we excluded outliers before calculating the mean absolute residual and standard deviation (gray *y*-axis error bars). Each S/N bin has a width of 1. Right: ratio of double vs. single component AIC parameters for the masked region of interested.

dependence on the S/N of the training set, or an effect from multiple line components in high S/N regions. In order to determine whether or not the S/N of the training set has a negative impact on high S/N regions, we create a set of 1000 synthetic data following the same prescription described before (Section 2); however, we allow the S/N to vary between 20 and 80 instead of stopping at 30. Because we only created 1000 synthetic spectra, we reduce the sampling rate of the velocity and broadening. This is not expected to have any effect on the results. We then apply our already trained network on the synthetic data. Figure A1 demonstrates that the network performs well for high S/N values. Thus the network is not biased for high S/N regions. Note that the S/N value used in this section is the true signal-to-noise ratio as compared to that used in Section 4.2 which is a proxy value calculated by dividing the Hα flux by its fit uncertainty.

In order to determine whether or not the regions of high S/N in the southwest field of M33 have single or double emission components, we turn to the standard ORCS fitting procedure. We chose a small region (2 × 2 pixels) in a high S/N region that also has a large broadening residual (*01:32:16.03, +30:48:00.71*). We selected pixels that fit the following prescription: have a broadening residual higher than 10 km s^{-1} and a signal-to-noise ratio over 12. We fit the Hα and N II doublet assuming a single emission component and a double emission component. The double emission fit resulted in a statistically significantly better fit statistic. This is a strong indication that the region is best described by a double emission component rather than a single emission component. Moreover, we computed the AIC parameter for each region defined by $\mathrm{AIC} = 2n - ln(L)$, where n is the number of fit parameters and L is the Gaussian likelihood function (e.g., Akaike 1987; Kieseppa 1997; Liddle 2007). In our case, the likelihood is Gaussian; therefore, the log-likelihood function reduces to the usual half χ-squared. The right-hand side of Figure A1 shows the ratio of the double component AIC parameter versus the single component AIC parameter defined as $\exp(-(\mathrm{AIC}_1 - \mathrm{AIC}_0)/2)$. Since the ratio is consistently below one, the double component model is favored over the single component model. We thus conclude that, at least in these regions, the rise in the residual value is due to the existence of double component emission. Therefore, we believe that Figure 10 does not reflect a failure of the network in high S/N regions, but rather a failure of the network in regions with double emission components that serendipitously appear in regions of high S/N in the southwest field of M33. Future work will explore the applicability of a modified network to estimate the broadening and velocity parameter in such regions.

ORCID iDs

Carter Rhea https://orcid.org/0000-0003-2001-1076
Laurie Rousseau-Nepton https://orcid.org/0000-0002-5136-6673
Simon Prunet https://orcid.org/0000-0002-1755-4582
Julie Hlavacek-Larrondo https://orcid.org/0000-0001-7271-7340
Sébastien Fabbro https://orcid.org/0000-0003-2239-7988

References

Abadi, M., Agarwal, A., Barham, P., et al. 2016, arXiv:1603.04467
Akaike, H. 1987, Psychometrika, 52, 317
Arsenault, R., & Roy, J.R. 1986, AJ, 92, 567
Bacon, R., Accardo, M., Adjali, L., et al. 2010, Proc. SPIE, 7735, 773508
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Baril, M., Grandmont, F., Mandar, J., et al. 2016, Proc. SPIE, 9908, 29
Baron, D. 2019, arXiv:1904.07248
Beck, R., Brandenburg, A., Moss, D., Shukurov, A., & Sokoloff, D. 1996, ARA&A, 34, 155
Bengio, Y., & Grandvalet, Y. 2004, JMLR, 5, 1089
Bertin, E. 1994, in Science with Astronomical Near-Infrared Sky Surveys, ed. N. Epchtein, A. Omont, B. Burton, & P. Persi (Dordrecht: Springer), 49
Bittner, A., Falcón-Barroso, J., Nedelchev, B., et al. 2019, A&A, 628
Bregman, J. N. 1980, ApJ, 236, 577
Bundy, K., Bershady, M. A., Law, D. R., et al. 2014, ApJ, 798, 7
Cawley, G. C., & Talbot, N. L. C. 2010, JMLR, 11, 2079
Cedrés, B., Beckman, J. E., Bongiovanni, A., et al. 2013, ApJL, 765, L24
Chollet, F. 2015, Keras, https://keras.io
Crawford, C. S., Allen, S. W., Ebeling, H., Edge, A. C., & Fabian, A. C. 1999, MNRAS, 306, 857
de Oliveira, C. A., Birkmann, S. M., Boeker, T., et al. 2018, arXiv:1805.06922 [astro-ph]
Dressler, A., Sandage, A., & Wilson, M. 1983, ApJL, 265, 664

Drissen, L., Martin, T., Rousseau-Nepton, L., et al. 2019, MNRAS, 485, 3930
Drissen, L., Rousseau-Nepton, L., Lavoie, S., et al. 2014, AdAst, 2014, 9
Epinat, B., Amram, P., & Marcelin, M. 2008, MNRAS, 390, 466
Fabbro, S., Venn, K., O'Briain, T., et al. 2018, MNRAS, 475, 2978
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, arXiv:1705.10877
Franco, J., Kurtz, S. E., García-Segura, G., & Hofner, P. 2000, Ap&SS, 272, 169
French, H. B. 1980, ApJ, 240, 41
Garnett, D. R., & Shields, G. A. 1987, ApJ, 317, 82
Gendron-Marsolais, M., Hlavacek-Larrondo, J., Martin, T. B., et al. 2018, MNRAS, 479, 28
Gil de Paz, A., Carrasco, E., Gallego, J., et al. 2012, Proc. SPIE, 8446, 84464Q
Haverkorn, M., Akahori, T., Carretti, E., et al. 2015, arXiv:1501.00416
Horowitz, L. 1974, ITASS, 22, ASSP-22
Hunter, J. D. 2007, CSE, 9, 90
Iwasaki, H., Ichinohe, Y., & Uchiyama, Y. 2019, MNRAS, 488, 4106
Kashino, D., & Inoue, A. K. 2019, MNRAS, 486, 1053
Kennicutt, R., & Oey, M. 1993, RMxAA, 27, 21
Keown, J., Francesco, J. D., Teimoorinia, H., Rosolowsky, E., & Chen, M. C.-Y. 2019, ApJ, 885, 32
Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, MNRAS, 372, 961
Kieseppa, I. 1997, The British Journal for the Philosophy of Science, 48, 21
Le Coarer, E., Rosado, M., Georgelin, Y., Viale, A., & Goldes, G. 1993, A&A, 280, 365
Leroy, A. K., Hughes, A., Schruba, A., et al. 2016, ApJ, 831, 16
Liddle, A. R. 2007, MNRAS, 377, L74
Martin, T., & Drissen, L. 2017, arXiv:1706.03230
Martin, T. B., Prunet, S., & Drissen, L. 2016, MNRAS, 463, 4223
Martins, F., Pomarès, M., Deharveng, L., Zavagno, A., & Bouret, J. C. 2010, A&A, 510, A32
McKinney, W. 2010, in Proc. 9th Python in Science Conf., 56
McLeod, A. F., Kruijssen, J. M. D., Weisz, D. R., et al. 2020, arXiv:1910.11270
Molinaro, A. M., Simon, R., & Pfeiffer, R. M. 2005, Bioinformatics, 21, 3301
Morisset, C., Delgado-Inglada, G., & Flores-Fajardo, N. 2015, RMxAA, 19, 103
Ntampaka, M., Trac, H., Sutherland, D. J., et al. 2016, ApJ, 831, 135
Ntampaka, M., ZuHone, J., Eisenstein, D., et al. 2019, ApJ, 876, 82
Odell, C. R. 1986, ApJ, 304, 767
Oey, M., & Kennicutt, R. 1993, ApJ, 411, 137
Olney, R., Kounkel, M., Schillinger, C., et al. 2020, arXiv:2002.08390
Osterbrock, D., & Ferland, G. 1989, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei (1st edn; Sausalito CA, USA: University Science Books)
Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019, A&A, 621, A26
Pavel, M. D., & Clemens, D. P. 2012, ApJ, 760, 150
Pérez-Montero, E., García-Benito, R., & Vílchez, J. M. 2019, MNRAS, 483, 3322
Picard, R. R., & Cook, R. D. 1984, Journal of the American Statistical Association, 79, 575
Price-Whelan, A. M., Sipőcz, B. M., Rix, H.-W., et al. 2018, AJ, 156, 18
Puertas, S. D., Iglesias-Páramo, J., Vilchez, J. M., et al. 2019, A&A, 629, A102
Quireza, C., Rood, R. T., Bania, T. M., Balser, D. S., & Maciel, W. J. 2006, ApJ, 653, 1226
Ramachandran, V., Hamann, W.-R., Hainich, R., et al. 2018, A&A, 615, A40
Ramachandran, V., Hamann, W.-R., Oskinova, L. M., et al. 2019, A&A, 625, A104
Relaño, M., & Beckman, J. E. 2005, A&A, 430, 911
Relaño, M., Beckman, J. E., Zurita, A., Rozas, M., & Giammanco, C. 2005, A&A, 431, 235
Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, A&A, 558, A33
Rousseau-Nepton, L., Martin, R. P., Robert, C., et al. 2019, MNRAS, 489, 5530
Rousseau-Nepton, L., Robert, C., Drissen, L., Martin, R. P., & Martin, T. 2018, MNRAS, 477, 4152
Rozas, M., Richer, M. G., Steffen, W., García-Segura, G., & López, J. A. 2007, A&A, 467, 603
Sadaghiani, M., Sanchez-Monge, A., Schilke, P., et al. 2019, arXiv:1911.06579
Sánchez, S. F., Pérez, E., Sánchez-Blázquez, P., et al. 2016, RMxAA, 52, 171
Sánchez, S. F., Rosales-Ortega, F. F., Marino, R. A., et al. 2012, A&A, 546, A2
Sancisi, R., Fraternali, F., Oosterloo, T., & van der Hulst, T. 2008, A&ARv, 15, 189
Scargle, J. D. 1982, ApJ, 263, 835
Schulz, M., & Stattegger, K. 1997, CG, 23, 929
Sharples, R., Bender, R., Agudo Berbel, A., et al. 2013, Msngr, 151, 21
Shields, G. A. 1990, ARA&A, 28, 525
Shields, G. A., & Tinsley, B. M. 1976, ApJ, 203, 66
Simonyan, K., Vedaldi, A., & Zisserman, A. 2013, arXiv:1312.6034
SOFUE, Y. 1995, PASJ, 47, 527
Stasińska, G., Izotov, Y., Morisset, C., & Guseva, N. 2015, A&A, 576, A83
Storrie-Lombardi, M. C., Lahav, O., Sodre, L., & Storrie-Lombardi, L. J. 1992, MNRAS, 259, 8P
Tetko, I., & Villa, A. 1997, NN, 10, 1361
Ucci, G., Ferrara, A., Gallerani, S., et al. 2019, MNRAS, 483, 1295
Ucci, G., Ferrara, A., Gallerani, S., & Pallottini, A. 2017, MNRAS, 465, 1144
Ucci, G., Ferrara, A., Pallottini, A., & Gallerani, S. 2018, MNRAS, 477, 1484
Vale Asari, N., Stasińska, G., Morisset, C., & Fernandes, R. C. 2016, MNRAS, 460, 1739
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Van Rossum, G., & Drake, F. L. 2009, Python 3 Reference Manual (Scott's Valley, CA: Create Soace)
Vasiliev, E. O., Moiseev, A. V., & Shchekinov, Y. A. 2015, BaltA, 24, 213
Veilleux, S., & Osterbrock, D. E. 1987, ApJS, 63, 295
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
Waskom, M., Botvinnik, O., O'Kane, D., et al. 2017, mwaskom/seaborn, v0.8.1, Zenodo, 10.5281/zenodo.883859
Weedman, D. W., Feldman, F. R., Balzano, V. A., et al. 1981, ApJ, 248, 105
Zeidler, P., Nota, A., Sabbi, E., et al. 2019, AJ, 158, 201
Zinchenko, I. A., Dors, O. L., Hagele, G. F., Cardaci, M. V., & Krabbe, A. C. 2019, MNRAS, 483, 1901