

A Machine Learning Approach to Integral Field Unit Spectroscopy Observations: III. Disentangling Multiple Components in H ii regions

Carter Rhea, Laurie Rousseau-Nepton, Simon Prunet, Julie Hlavacek-Larrondo, R. Pierre Pierre Martin, Kathryn Grasha, Natalia Vale Asari, Théophile Bégin, Benjamin Vigneron, Myriam Prasow-Émond

▶ **To cite this version:**

Carter Rhea, Laurie Rousseau-Nepton, Simon Prunet, Julie Hlavacek-Larrondo, R. Pierre Pierre Martin, et al.. A Machine Learning Approach to Integral Field Unit Spectroscopy Observations: III. Disentangling Multiple Components in H ii regions. The Astrophysical Journal, 2021, 923 (2), pp.169. 10.3847/1538-4357/ac2c66 . hal-03534347

HAL Id: hal-03534347
https://cnrs.hal.science/hal-03534347

Submitted on 19 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Draft version October 4, 2021
Typeset using LaTeX **twocolumn** style in AASTeX63

arXiv:2110.00569v1 [astro-ph.GA] 1 Oct 2021

A Machine Learning Approach to Integral Field Unit Spectroscopy Observations: III. Disentangling Multiple Components in H II regions

Carter L. Rhea [1,2] Laurie Rousseau-Nepton [3] Simon Prunet [3,4] Julie Hlavacek-Larrondo [5] R. Pierre Martin [6] Kathryn Grasha [7,8] Natalia Vale Asari [9,10,11] Théophile Bégin [1] Benjamin Vigneron [1] and Myriam Prasow-Émond [1]

[1] *Département de Physique, Université de Montréal, Succ. Centre-Ville, Montréal, Québec, H3C 3J7, Canada*
[2] *Centre de Recherche en Astrophysique du Québec (CRAQ), Québec, QC, G1V 0A6, Canada*
[3] *Canada-France-Hawaii Telescope, Kamuela, HI, United States*
[4] *Laboratoire Lagrange, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Parc Valrose, 06104 Nice Cedex 2, France*
[5] *Département de Physique, Université de Montréal, Succ. Centre-Ville, Montréal, uébec, H3C 3J7, Canada*
[6] *Department of Physics and Astronomy, University of Hawaii at Hilo, 200 W. Kawili St., Hilo, 96720, USA*
[7] *Research School of Astronomy and Astrophysics, Australian National University, Weston Creek, ACT 2611, Australia*
[8] *ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*
[9] *Departamento de Física–CFM, Universidade Federal de Santa Catarina, C.P. 476, 88040-900, Florianópolis, SC, Brazil*
[10] *School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK*
[11] *Royal Society–Newton Advanced Fellowship*

(Received; Revised; Accepted October 4, 2021)

Submitted to ApJ

ABSTRACT

In the first two papers of this series (Rhea et al. 2020b; Rhea et al. 2021), we demonstrated the dynamism of machine learning applied to optical spectral analysis by using neural networks to extract kinematic parameters and emission-line ratios directly from the spectra observed by the SITELLE instrument located at the Canada-France-Hawai'i Telescope. In this third installment, we develop a framework using a convolutional neural network trained on synthetic spectra to determine the number of line-of-sight components present in the SN3 filter (656–683nm) spectral range of SITELLE. We compare this methodology to standard practice using Bayesian Inference. Our results demonstrate that a neural network approach returns more accurate results and uses less computational resources over a range of spectral resolutions. Furthermore, we apply the network to SITELLE observations of the merging galaxy system NGC2207/IC2163. We find that the closest interacting sector and the central regions of the galaxies are best characterized by two line-of-sight components while the outskirts and spiral arms are well-constrained by a single component. Determining the number of resolvable components is crucial in disentangling different galactic components in merging systems and properly extracting their respective kinematics.

Keywords: Machine Learning; ISM; Galaxies; Resolved Emission Components

1. INTRODUCTION

Systems of merging galaxies are critical tracers of the evolutionary history of the universe according to the well accepted hierarchical model of galaxy formation (e.g. Conselice 2014). Simulations and observations of interacting systems and galaxy clusters reveal that the merging process is responsible for the creation of a significant fraction of stars as well as modifying the distribution of gas in galaxies and in the intergalactic medium of galaxy clusters (e.g. Liu et al. 2016; Barnes 2002; Struck 2007); the mergers are additionally responsible for starbursts occuring after the merger (known as post-merger starbursts) which are phases of intense star formation (e.g. Wild et al. 2009; Trouille et al. 2013). The rate and

Corresponding author: Carter L. Rhea
carter.rhea@umontreal.ca

efficiency of the star formation activity vary in time all throughout the merger process and depends on a wide variety of collision parameters but is most importantly related to the galaxies' initial gas content (e.g. Barnes & Hernquist 1996; Darg et al. 2010; Donzelli & Pastoriza 2000).

The evolution and kinematics of star forming gas in merging systems can be observed in the optical and near infrared wavelengths. Specifically, several optical lines are used as tracers of ionized gas associated with star-forming regions and other emission line objects: H $\alpha\lambda$6563, and H $\beta\lambda$4861, ionized oxygen (i.e. [O II]$\lambda\lambda$3726, 3729, and [O III]$\lambda\lambda$4959, 5007), ionized nitrogen (i.e. [N II]$\lambda\lambda$6548, 6583), and ionized sulfur (i.e. [S II]$\lambda\lambda$6716, 6731) (e.g. Osterbrock & Ferland 1989; Shields 1990; Veilleux & Osterbrock 1987; the lines' wavelengths are quoted in Angstroms). These strong emission lines help identify excitation mechanisms and chemical abundances present in the gas (Baldwin et al. 1981; Crawford et al. 1999; Kewley et al. 2001).

To measure the emission lines' flux within a spectrum, it is essential to fit the lines using the proper model with parameters that allow to represent the intensities in all spectral elements. Estimating the number of components of each underlying emission lines (i.e. multiple lines with multiple velocity components as would be expected from merging galaxy systems) is also crucial to extract meaningful information from the lines. Standard methodologies require fitting both single and double component models and computing their Bayes factor, or approximating it using a proxy such as the Akaike (or Bayesian) Information Criterion (AIC; Rhea et al. 2020b). However, these methods are highly reliant on the accuracy of the fits (e.g. Kieseppä 1997; Pooley & Marion 2018). Several promising new methods have been proposed utilizing machine learning algorithms to solve this problem (Hampton et al. 2017; Keown et al. 2019; Rhea et al. 2020a). In particular, CLOVER, developed in Keown et al. (2019), uses a convolutional neural network to classify high-resolution radio emission lines as having either single or double underlying components. In this paper, we expand upon this methodology for medium resolution, Fourier Transform IFU spectra taken by the SITELLE (*Spectromètre Imageur à Transformée de Fourier pour l'Etude en Long et en Large de raies d'Emission*) instrument at the Canada-France-Hawai'i Telescope (Baril et al. 2016). Each SITELLE observation contains approximately 4 million pixels and, thus, produces 4 million spectra of a given resolution set by the principal investigator ($1 < R < 10,000$; Drissen et al. 2014; Drissen et al. 2019; Martin & Drissen 2017). The instrumental line function of SITELLE is described as a sinc model convolved with a Gaussian to represent intrinsic line broadening which requires special care during the fitting process (Martin et al. 2016). Therefore, the development of machine learning applications for SITELLE demands special treatment of the underlying emission profiles.

In this paper, we explore using a convolutional neural network to classify extragalactic nebulae imaged with SITELLE as having either a single underlying velocity component or two. In § 2, we develop a synthetic set of data, explain the conventional methods for determining the number of underlying components, and develop a convolutional neural network architecture. In § 3, we discuss the results of classifying the test set data using the AIC, Bayesian inference, and our convolutional neural network (CNN). We compare the three methodologies, discuss applicability to other instruments, and test out the algorithm on real SITELLE observations of the merging galaxy system NGC2207/IC2163 in §4. Concluding remarks are made in §5.

2. METHODOLOGY

In the first two articles of the series, Rhea et al. (2020b) & Rhea et al. (2021), we explored the application of convolutional and artificial neural networks to calculate the kinematic parameters and flux ratios from SITELLE spectra. In this paper, we employ a convolutional neural network to determine the appropriate number of underlying emission components in [S II]λ6717, [S II]λ6731, [N II]λ6584, Hα(6563Å), and [N II]λ6548. To train the algorithm, we start by developing a set of synthetic SITELLE spectra.

2.1. *Synthetic Data*

The synthetic data is created using the `ORB` – *Outil de Reduction Binoculaire* – software package described in Martin et al. (2016) and Martin et al. (2012). The `create_cm1_lines_model` is used to create the following emission lines: [S II]λ6717, [S II]λ6731, [N II]λ6584, Hα(6563Å), and [N II]λ6548. We use these lines since they are all present in the SN3 (651-685 nm) filter. `ORB` creates synthetic spectra by taking the spectral resolution, line amplitude, line function, line broadening, and velocity as user inputs. As in other papers, these spectra are representative of those expected from the SIGNALS (*Star formation, Ionized Gas, and Nebular Abundances Legacy Survey*; Rousseau-Nepton et al. 2019). We randomly sample the resolution from a uniform distribution varying between R=4800 and R=5000; this variation mimics expected resolution variations within the SITELLE field of view, that are observed in SITELLE data cubes.

The velocity value is randomly chosen from a uniform distribution between $-500\,\mathrm{km\,s^{-1}}$ and $500\,\mathrm{km\,s^{-1}}$; these values well encompass the typical rotation velocity range for galaxies found in the SIGNALS program. Additionally, we select the broadening from an uniform distribution between $5\,\mathrm{km\,s^{-1}}$ and $20\,\mathrm{km\,s^{-1}}$. We selected the lower limit since SITELLE cannot resolve broadening beneath this level (Rousseau-Nepton et al. 2019). We note that shocks in the ISM are not being considered here – they would require an increase in the maximum broadening. We verified that the chosen lower limit does not impact the classification process by testing different lower bounds ($3\,\mathrm{km\,s^{-1}}$ and $1\,\mathrm{km\,s^{-1}}$). We selected $20\,\mathrm{km\,s^{-1}}$ since it is the expected upper limit of dispersion in HII regions based onon previous experience in SITELLE analysis (Rhea et al. 2020b). Moreover, we randomly vary the signal-to-noise ratio concerning the Hα emission between 5 and 30. We propagate the noise to the other emission lines by adding it to each spectral channel. The noise factor itself is sampled from Gaussian with a sigma of 1. Therefore, each spectrum has a different signal-to-noise ratio described by a Gaussian centered between 5 and 30 with a sigma of 1. The last feature required to create synthetic spectra is the relative amplitude of the strong emission lines. Following the methodology described in detail in Rhea et al. (2020b), relative line amplitude of the HII regions were sampled from the Mexican Million Models database (3MdB; Morisset et al. 2015) BOND simulations (Asari et al. 2016).

We created a set of 10,000 spectra with a single component; therefore, there are five emission lines present ([S II]λ6717, [S II]λ6731, [N II]λ6584, Hα, and [N II]λ6548) with the same velocity and broadening values. An additional 10,000 spectra with two components was also created; therefore, there are two sets of five emission lines present with the same velocity and broadening values within each set. Each set has a different velocity and broadening value. The amplitudes for each set come from different instances of the `BOND` simulations from `3MdB`. Thus, we have 20,000 spectra in total. Following standard procedure, we use 70% of the synthetic data for the training set, 20% for the validation set, and 10% for the test set (e.g., Breiman 2001). We note that individual lines may have different kinematics; however, we do not consider this in our analysis for simplicity's sake.

2.2. *Bayesian Inference*

In the past decade, Bayesian statistics have become house-hold tools for astronomers; Bayesian techniques are used to study eclipsing exoplanet signals (Taaki et al. 2020; Ruffio et al. 2018), fitting spectra in order to extract model parameters (Sereno 2016; Sharma 2017), and model comparison (Jenkins & Peacock 2011; Trotta 2007). In this section, we outline the mathematics behind Bayesian inference, how it can be used to compare models, and its practical implementations in `python`.

Bayesian inference is the process of uncovering underlying parameter distributions through the use of Bayes' theorem. Bayes' theorem states

$$P(\Theta_M|Y,I) = \frac{P(Y|\Theta_M,I)\times P(\Theta_M|I)}{P(Y|I)}, \tag{1}$$

where Θ_M describes a set of model parameters for model M, Y is the set of observed data, and I represents any assumed information. The left-hand side of the equation is known as the posterior distribution, and it describes the probability distribution of model parameters given the existing data. On the right-hand side, we have the likelihood distribution, $P(Y|\Theta_M,I)$, which describes the distribution of the data given a set of model parameter values, the prior distribution, $P(\Theta_M,I)$, describing our prior knowledge or believes regarding the distribution of the model parameters, and the evidence, $P(Y|I)$ or Z. In many scenarios, Bayes' theorem is rewritten as

$$P(\Theta_M|Y,I) \propto P(Y|\Theta_M,I)\times P(\Theta_M|I), \tag{2}$$

where the evidence is ignored. However, in the case of model comparison, we wish to calculate the Bayes factor, which is defined as the ratio of the evidences calculated using two distinct models. Therefore, the evidence must be calculated; note that the evidence is defined as the integral over all the likelihood multiplied by the prior model parameters.

$$Z = \int_{\theta_M\in\Theta_M} P(Y|\Theta_M,I)\times P(\Theta_M|I)d\theta_m. \tag{3}$$

We write the Bayes' factor, $P(M|D)$, as

$$P(M|D) = \frac{Z_1}{Z_2} \tag{4}$$

We adopt a standard Gaussian likelihood function of the following form:

$$P(Y|M_\Theta,I) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(\frac{-(Y-Y_M(M_\Theta))^2}{2\sigma_0^2}\right) \tag{5}$$

where $Y_M(M_\Theta)$ is the assumed model evaluated given the model parameters M_Θ and σ_0 are the measurement errors. For the sake of the calculations simplicity (discussed in §4.1), we employ a standard Gaussian to model the emission line. The Gaussian has the following form:

$$Y_M(M_\Theta) = Y_M(A,x_0,\sigma) = A\exp\left(\frac{-(x-x_0)^2}{2\sigma^2}\right) \tag{6}$$

where A is the amplitude of the Gaussian, x is the wavelength channel, x_0 is the position of the line, and σ is the width of the line. We emphasize that we calculate the posteriors for the parameters $M_\Theta = [A, x_0, \sigma]$ in the case of a single Gaussian model. We adopt linearly uniform priors for all the fit parameters.

Unfortunately, the calculation of the evidence is exceptionally costly in high-dimensional spaces, and alternative methods to calculate the evidence are required. Nested sampling is one of the most popular methods for accurately and rapidly calculating the evidence (e.g. Skilling 2006; Chopin & Robert 2010). We, therefore, use the nested sampling code `dynesty` for our calculations (Speagle 2020). Since the calculations are computationally expensive, we explore a faster algorithm – a convolutional neural network.

2.3. *Convolutional Neural Networks*

Neural networks, and their extension convolutional neural networks, have been used extensively to solve astrophysical problems (e.g. Baron 2019; Davies et al. 2019; Aniyan & Thorat 2017; Kim & Brunner 2017). Recently, convolutional neural networks have been shown to be efficient at rapidly and accurately extracting parameters from spectra (e.g. Fabbro et al. 2018; O'Briain et al. 2021;Keown et al. 2019; Rhea et al. 2020b). For a detailed review of convolutional neural networks, we direct the reader towards the following works: Kuo (2016), Liu et al. (2016), Khan et al. (2020). In particular, by interpreting model selection as a classification problem, neural networks with a softmax activation on the last layer can be used as a substitute to the reference Bayesian model selection approach, with a substantial gain in computational complexity.

For this work, we adopt the structure of the network STARNET developed in Fabbro et al. (2018) and used in our previous work (Rhea et al. 2020b). We note that the hyperparameters (i.e. filters) are tuned separately and are thus different from those used in Fabbro et al. (2018). The network is as follows:

1. Convolutional layer with 4 filters of 16 elements activated with the `relu` function
2. Convolutional layer with 4 filters of 8 elements activated with the `relu` function
3. Pooling layer with a size of 8 elements
4. Flattening Layer
5. 20% Dropout
6. Fully connected layer with 1000 nodes activated with the `relu` function and regularized using ℓ_2 regularization
7. 20% Dropout
8. Fully connected layer with 1000 nodes activated with the `relu` function and regularized using ℓ_2 regularization
9. Fully connected output layer with two nodes activated with the `softmax` function

Hyper Parameter	Value
Batch Number	2
ℓ_2 Regularization	$5e^{-5}$
Optimizer	Adam
Kernel Initializer	Glorot Uniform
Max Epochs	25

We note that the use of the `softmax` activation function in the output layer not only generates a binary classification but also assigns a probability of accuracy to the classification.

Although we adopt the convolutional neural network, STARNET, described in detail in Fabbro et al. 2018, we apply a complete suite of hyper-parameter tuning. Using the `sklearn` grid search implementation, we optimized the number of convolutional kernels in each layer (4 in both layers) and the length of each filter (16 elements in the first layer and 8 elements in the second layer. We also optimized the number of spectra fed into the network at a time (i.e., the optimal batch size; we found this to be 2 spectra), the best optimization algorithm, `Adam`, the length of the convolutional pooling in each layer, 8. These results are summarized in table 2.3.

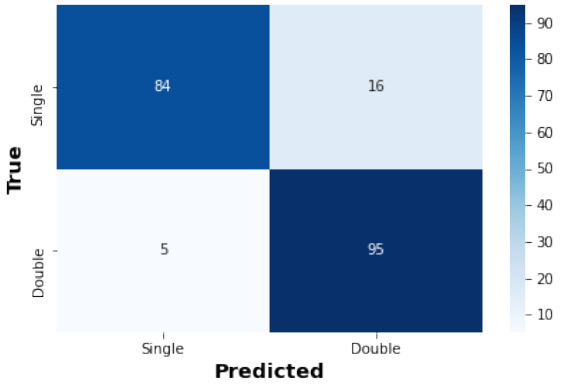
Since the resolution is free to vary uniformly from 4800 to 5000 and the network requires inputs of a constant size, we interpolate the spectra using linear cardinal splines using a grid taken from a reference spectrum with a resolution set to 5000. This does not affect the fidelity of the spectrum.

3. RESULTS

In this section, we discuss the results of the Bayesian inference and CNN algorithms. We use confusion matrices to compare the efficiencies of the different methodologies. Confusion matrices compare the true categorizations (Y-axis) with the predicted categorizations (X-axis). A perfectly diagonal confusion matrix indicates that the method accurately predicts the number of line-of-sight components 100% of the time. Cross terms represent misclassifications. The values quoted are percentages.

3.1. *Bayesian Inference*

Figure 1. Confusion matrix for the Bayesian inference calculations.



Following the procedure described in §2.2, we fit a single Gaussian model and a double Gaussian model using the `dynesty` implementation of static nested sampling to the same test set used in the CNN calculations. We fit only the Hα line in order to reduce the computational cost (see §4.1 for a more detailed discussion on the repercussions of this decision). We then used the reported outputs to calculate the Bayes' factor. The double component model is considered favorable if the Bayes' factor is greater than 1.08 (this value was experimentally determined to return the best classifications[1]); otherwise, a single component model is favorable. We used the static nested sampling implementation over the more accurate dynamic nested sampling implementation since the latter increased the computational costs by more than an order of magnitude – this is further discussed in §4. The results are shown in figure 1. We investigated the cases in which a double component model was mistaken for a single component model (approximately 15% of the time) and discovered that the majority of misclassifications occurred when the absolute velocity difference between the two components was less than 250 $\mathrm{km\,s^{-1}}$ (see the next subsection for further discussion).

3.2. *Convolutional Neural Network*

The convolutional neural network was trained and validated using 70% and 20% of the synthetic data. At the end of the process, the overall accuracy of the algorithm was 93%. The mse (mean squared error) on both the test and validation set is 0.44; this indicates that the network is not over-fitting. Over-fitting occurs when the network learns the test set too well and cannot generalize to other data (i.e., the validation or test set). The confusion matrix for the test set (see figure 2) reveals near-perfect precision in classifying single line-of-sight component spectra. Moreover, the matrix demonstrates that the network classifies double line-of-sight component spectra accurately 90% of the time; the remaining 10% are misclassified as single components. We note that this is less than the 95% accuracy of the Bayesian methodology.

In order to establish that the network is not over-fitting, we applied the standard k-fold cross-validation with $k = 10$. Additionally, we applied a modified k-fold cross-validation algorithm in which only the training and validation set are varied while the test set remains consistent across all folds. The reported accuracy of the model for each fold regardless of implementation remained constant (less than 5% variation). This is a strong indication that the network is not over-fitting, and, thus, it can be generalized to unseen data.

Moreover, we examined the regimes in which the network fails to predict the number of underlying components accurately. In figure 3, we plot a histogram of the absolute velocity difference for the test set spectra with double line-of-sight components. The stacked histogram shows that the incorrectly categorized spectra (rose) are primarily clustered around absolute velocity differences less than 250 $\mathrm{km\,s^{-1}}$. This trend is expected since, at these low absolute velocity differences, the components are blended considerably. We show four spectra illustrating this issue in appendix A.

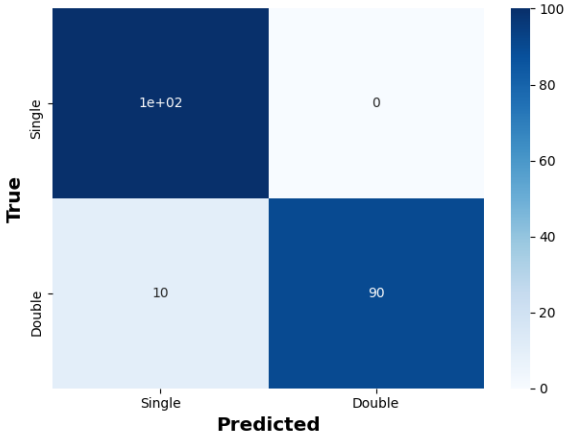


Figure 2. Confusion matrix for the convolutional neural network described in §2.3.

3.3. *Other Machine Learning Algorithms*

In addition to using a convolutional neural network to estimate the number of emission-line components,

[1] In order to determine the cutoff Bayes factor we varied the cutoff from 1 to 5 in increments of 0.05 and calculated the f1-score given the chosen factor. We then took the Bayes factor that returned the best f1-score.

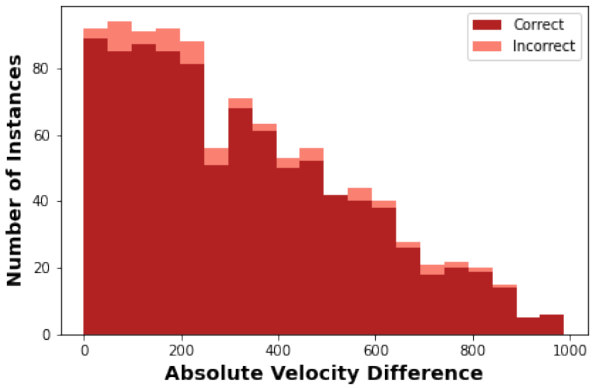


Figure 3. Absolute velocity difference vs number of instances for mis-categorized spectra containing two underlying components.

we tested another common machine learning algorithm for binary classification problems: the random forest. While random forests are themselves complex machine learning algorithms, we do not discuss them in detail here since the purpose of this section is to demonstrate the efficacy of other machine learning algorithms to solve our problem; instead we point the reader to (Breiman 2001) and the references therein. Although we test several configurations for the random forest classifier, our final classifier has a total of 50 independent estimators (decision trees) with a maximum recursion depth of 5 levels. The split criterion is based on entropy calculations instead of gini calculations. As evidenced by figure 4, the random forest algorithm does not achieve the same fidelity in categorization as the neural network.

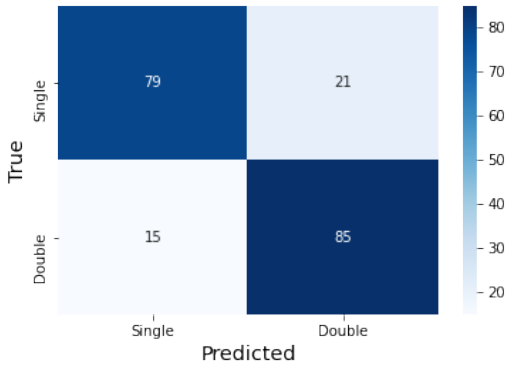


Figure 4. Confusion matrix for the random forest classifier tested in this work

4. DISCUSSION

4.1. *Comparison of Methods*

As evidenced by the confusion matrices shown in figures 1 and 2, the convolutional neural network outperforms the Bayesian inference classifications in the single component case. Notably, the CNN slightly (3%) under-performs the Bayesian Inference analysis for the classification of double line-of-sight components. Since the two methods were tested on the same set of data (the CNN's test set), the methodologies' comparison is unbiased with regards to the data they are tested on. Although the Bayesian inference methodology returns similar results to that of the CNN, the method takes two orders of magnitude more computational time, rendering it impractical. The Bayesian method takes approximately 200 hours to fit the test set, while the network takes 45 minutes to train and approximately 1 second to predict the test set. Additionally, the creation of the synthetic data set takes approximately 20 minutes. Therefore, the total time for the CNN is slightly over 1 hour.

Generally, Bayesian Inference sets a baseline for fitting results; however, our results indicate that the CNN is outperforming the Bayesian inference model estimates for the single line-of-sight component scenario since it is more accurate for single line-of-sight component classifications. We attribute this discrepancy to the chosen calculation of the Bayesian evidence and the model used. In order to reduce the number of parameters to a more computationally manageable size, we only fit a Gaussian (or two Gaussians) to the Hα component. In doing so, we must marginalize over a 3 or 6 dimensional space, respectively (there are 3 components used to describe each Gaussian in the Bayesian approach). If we were to fit all five lines, we would be required to marginalize over a 15 or 30 dimensional space thus further increasing the required computational time. Therefore, we emphasize that the CNN is outperforming the Bayesian approach only because the Bayesian approach is incomplete since it does not fully treat the ILS of SITELLE nor does it take into account every line present in the SN3 spectra. We further note that a Bayesian approach taking into account all lines and the proper ILS is not implemented since it is too computationally expensive at this time.

We also note the work of González-Gaitán et al. (2019-01-21) in which the authors use spatial information and prior estimations on the spatial correlations to drastically reduce the required computational time using Bayesian methods. However, a direct comparison between our works is beyond the scope of this paper.

4.2. *Differing Resolution*

In order to port this methodology to observations that are not part of the SIGNALS collaboration, we create

an alternative test set at a different resolution. We select a resolution of 3000 to test the algorithm's efficacy on lower-resolution spectra. After creating a set of synthetic data following the same procedure described in section §2.1 with $R \sim 3000$, we train, validate, and test the algorithm on this lower-resolution set. Figure 5 demonstrates that the categorizations' fidelity does not change for single line-of-sight component spectra. At the same time, the accuracy is reduced by 10% for double line-of-sight component spectra. Additionally, we followed the same procedure for $R \sim 1000$ spectra; however, the network fails to achieve the same level of efficacy due to the shallow resolution. We, therefore, do not suggest using the network on spectra with a resolution considerably lower than $R \sim 3000$.

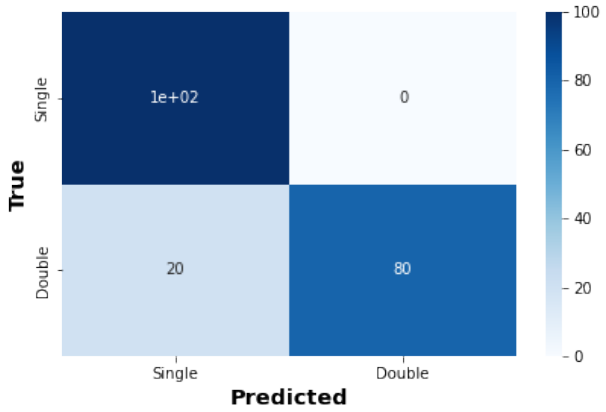


Figure 5. Confusion matrix for the convolutional neural network applied to spectra with $R \sim 3000$. The rows sum to 100 (unless there is a discrepancy due to rounding).

4.3. *Application to More Components*

Although we have only considered the problem as a binary classification (i.e. either single or double component) due to the constraints of the SIGNALS observing campaign, it is possible that a spectrum can contain more than two components. Therefore, we constructed a set of 1,000 synthetic spectra comprised of three underlying emission components following the same methodology as outlined in §2.1. We then apply our network to the spectra. Figure 6 demonstrates that the network classifies the majority (∼95%) of the three-component spectra as having two components. Therefore, we suggest the following interpretation if a region is suspected to have more than two components: a classification as two components should be considered as *at least* two components. These results are similar to those found in Rhea et al. 2020a.

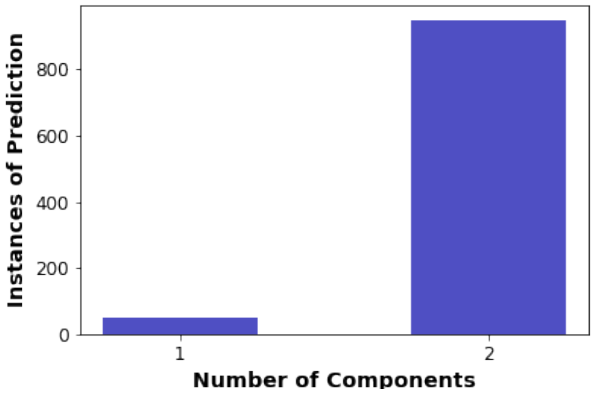


Figure 6. Histogram portraying the number of instances the network classified a spectrum containing three underlying emission components as only have one or two components. There were a total of 51 spectra classified as having only a single component while the remaining 949 spectra were classified as having two spectra.

4.4. *Application to Real SITELLE Observations*

In order to demonstrate the utility of this methodology, it is applied to actual SITELLE observations at R∼3000 of the interacting pair of spiral galaxies NGC 2207 and IC 2163 (P.I. R. Pierre Martin). This system (D = 35 Mpc) is undergoing a grazing collision and has been studied from X-rays to cm wavelengths (e.g. Kaufman et al. 2012; Elmegreen et al. 2017). The closest interaction of both galaxies is estimated to have happened 200–400 Myr ago (Kaufman et al. 2012) and, according to models, both galaxies are expected to merge in the future (Struck 2007). SITELLE observations cover the entire interaction, including zones not severely affected by the collision, tidal tails, and the zone of closest proximity between both galaxies. Therefore, this target presents an ideal test-bed for our algorithm. We note that no $R \sim 5000$ SITELLE observations of merging galaxies exist at the time of writing this paper.

In addition to applying the neural network to the SITELLE data cube, we obtain a deep image of the SITELLE cube using `ORCS`. The deep image is a 2d, stacked representation of the 3d data cube (i.e., the spectral information is compressed to a single pixel). The neural network returns a map of the galaxy where each pixel is designated as having either a single or double component. A comparison of these maps (the component map and the deep image) is shown in figure 7. The component map reveals several interesting features: double components in the bulges of each spiral galaxy, the double component emission in the current closest proximity zone (green circle), and single component emission in the spiral arms. Additionally, the com-

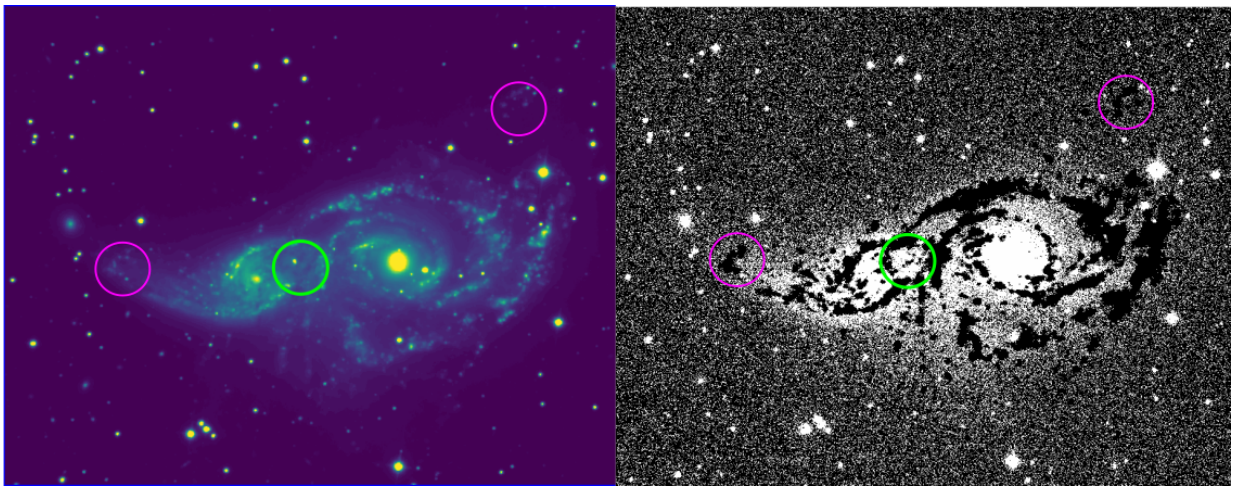


Figure 7. Left: Deep SITELLE image of the N2207/IC2163 system created with `ORCS`. This panel shows the stacked optical emission in the component galaxies. The colors (from blue to yellow) indicate the integrated flux in a spaxel. Several structures such as the spiral arms, bulges, tidal tails, and diffuse emission regions stand out. Right: Component map for the NGC2207/IC2163 system. White pixels correspond to double component emission. Black pixels correspond to single component emission. The green circle highlights the mixing region while the magenta circles highlight diffuse regions.

ponent map classifies diffuse regions (indicated in magenta circles) as single component emission regions. We note that the HII regions in the center of each galaxy are also categorized by double component emission. These findings are consistent with the literature (i.e. Soto et al. 2012). Since we use a `softmax` activation function in the final layer of the network, we also obtain probability maps. These maps in general can be used to make probability cuts reflecting regions for which the network is sure of its classification. We discuss this in more detail in the appendix *C*. We hypothesize that since the bulges are being classified with a high probability as having two components these regions have at least two components; this is consistent with our findings in §6. These results indicate that the network's classifications should be accepted cautiously for regions which are heavily contaminated by astrophysical processes such as AGN or a bright stellar component. In regions that correspond to the deep image background (i.e., not part of either galaxy in the system), the component map is noisy. This indicates that the component map is reliable only for regions of sufficient flux – which is expected. We stress that we do not expect to be able to resolve more than two emission-line components given the resolution of SIGNALS SN3 observations (R∼5000) and the distance to the objects. Expanding the network to be a multi-component classifier is a viable avenue for future research.

4.5. *Scientific Implications*

Determining the number of resolvable line-of-sight components in galaxy spectra carries several important scientific implications. In situations where two line of sight components are present and resolvable, it is important to fit both lines separately instead of treating the emission as a single component for several reasons. Distinguishing a region of multiple component line-of-sight emission from a large blended region will allow for the correct characterization of thermal broadening in H II regions which will help untangle the reaction of the gas to the ionizing source. Moreover, if the two components have considerably distinct kinematic parameters (velocity and broadening), treating the emission as a single component will lead to a misclassification of the kinematics of the region which can lead to incorrect interpretations of the star forming mechanisms present. In effect, a misunderstanding of the number of resolvable line-of-sight components present will lead to a misrepresentation of the underlying physics.

Additionally, mapping out regions in which two line-of-sight components is crucial in disentangling galaxy systems undergoing a merger. In doing so, the kinematics of the component galaxies can be studied; moreover, treating each galaxy's emission separately leads to more accurate calculations of the emission line fluxes. This is crucial for understanding the underlying ionization mechanisms at play in mixing regions of merger systems (e.g., Baldwin et al. 1981; Kewley et al. 2001; Kewley et al. 2019; Rich et al. 2015). Furthermore, by applying the networks developed in Rhea et al. (2020b) and Rhea et al. (2021) to the appropriate regions, kinematic parameters and emission-line ratios can be recovered. The authors note that those networks function only for a single component spectrum; future work will demonstrate their extension to double component spectra.

4.6. *Prospects of application in other instruments*

Although we focus on SITELLE data cubes in this article, the methodology described herein can be readily ported to any other high-resolution integral field unit that has access to the [N II] doublet, the [S II] doublet, and Hα emission lines such as the Multi Unit Spectroscopic Explorer, MUSE (e.g., Bacon et al. 2010; Kreckel et al. 2019; Kreckel et al. 2020; Foster et al. 2020). When applying this methodology to other instruments, it is essential to repeat the creation of synthetic data and network training, validation, and testing if the instrumental response function is not a `sincgauss`.

5. CONCLUSIONS

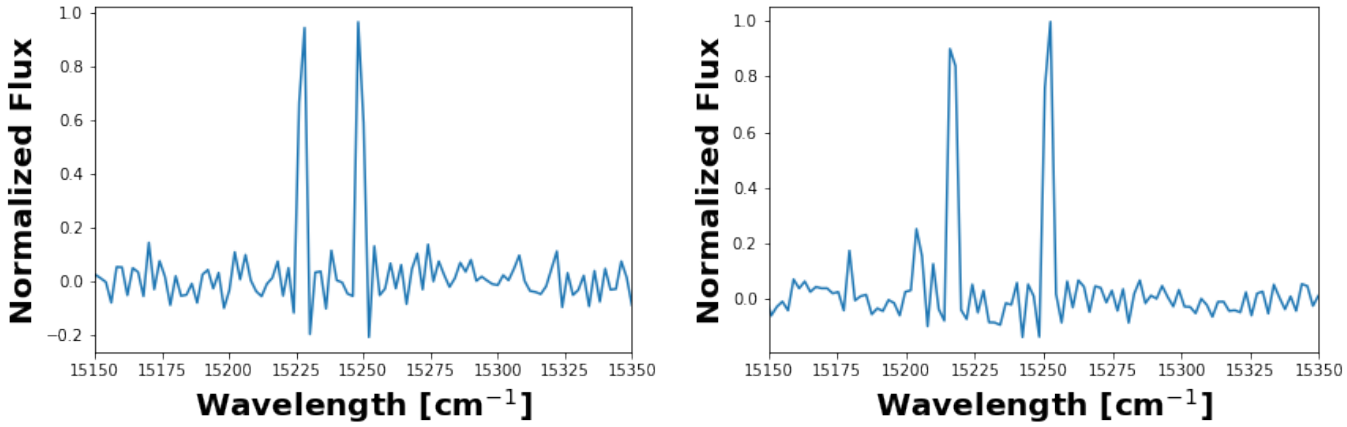
Machine learning algorithms present a novel approach for spectral analysis. In the first two papers of this series, we demonstrated the efficacy of convolutional and traditional neural networks at extracting kinematic parameters and emission-line flux ratios from SITELLE spectra (Rhea et al. 2020b; Rhea et al. 2021). In this paper, the third of the series, we develop a convolutional neural network to classify spectra as having either a single or double line-of-sight component. This systematic method will be critical for disentangling components in merger systems, HII regions, and supernova remnants. We demonstrate that the network outperforms Bayesian inference model comparisons in the single-component case and recovers similar accuracy in the double-component case. Moreover, the computational costs associated with the CNN training and subsequent application are several of orders of magnitude lower than the cost of the full Bayesian approach. In order to demonstrate the applicability of this network to real SITELLE data, we apply the network to actual SITELLE observations of the merging galaxy system NGC2207/IC2163. We find that the central regions of the individual galaxies and the closest proximity region are best categorized by double components. At the same time, the spiral arms and diffuse emission on the outskirts are best described by single component emission. The code can be found, along with code from the previous papers, at https://github.com/sitelle-signals/Pamplemousse.

ACKNOWLEDGMENTS

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.

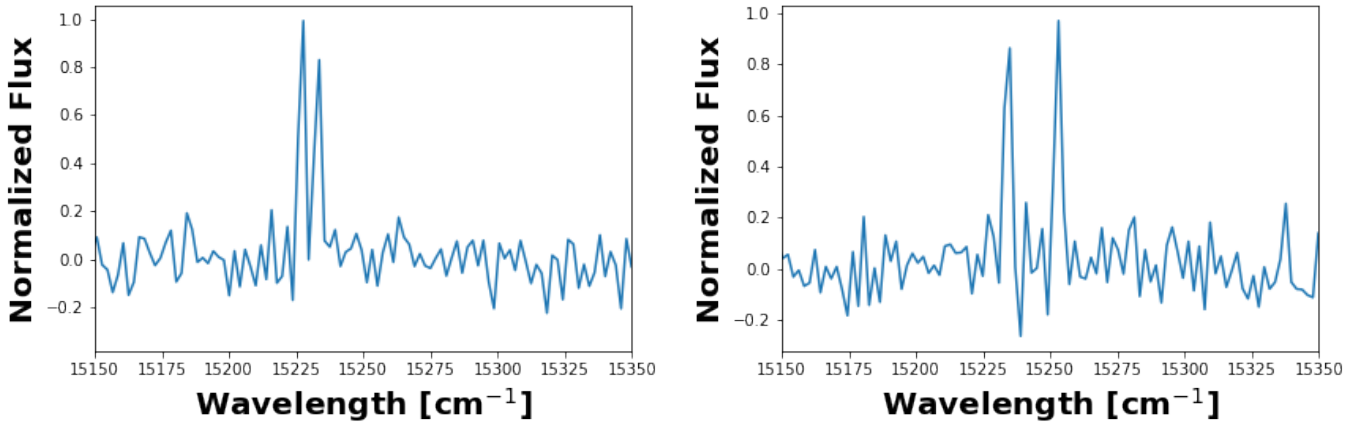
C. R. acknowledges financial support from the physics department of the Université de Montréal, IVADO, and the FRQNT. J. H.-L. acknowledges support from NSERC via the Discovery grant program, as well as the Canada Research Chair program. NVA acknowledges support of the Royal Society and the Newton Fund via the award of a Royal Society–Newton Advanced Fellowship (grant NAF\R1\180403), and of Fundação de Amparo á Pesquisa e Inovação de Santa Catarina (FAPESC) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We thank the anonymous referee for their insightful comments.

Software: python (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), (Robitaille et al. 2013), tensorflow (Abadi et al. 2015), keras (Chollet 2015) .



(a) Correctly classified double component spectrum with an absolute velocity difference of 85km s^{-1}

(b) Correctly classified double component spectrum with an absolute velocity difference of 500km s^{-1}



(c) Incorrectly classified double component spectrum with an absolute velocity difference of 146km s^{-1}

(d) Incorrectly classified double component spectrum with an absolute velocity difference of 465km s^{-1}

Figure 8.

APPENDIX

A. ILLUSTRATIVE SPECTRA

In this section we show four spectra which illlustrate correct classifications and misclassifications for double component spectra. We further divide the spectra by the absolute velocity difference of their components; we use 250 km s^{-1} as the divider. The top row images (a and b) are spectra with two components that were correctly categorized while the bottom row images (c and d) are spectra with two components that were incorrectly categorized. Similarly, the left images (a and c) contain components with an absolute velocity difference less than 250 km s^{-1} while the right images (b and d) contain components with an absolute velocity difference greater than 250 km s^{-1}.

B. CORNER PLOTS FROM DYNESTY

In this section we show the corner plots for a spectrum with two emission-line components assuming a likelihood with only a single Gaussian and a likelihood with two Gaussians.

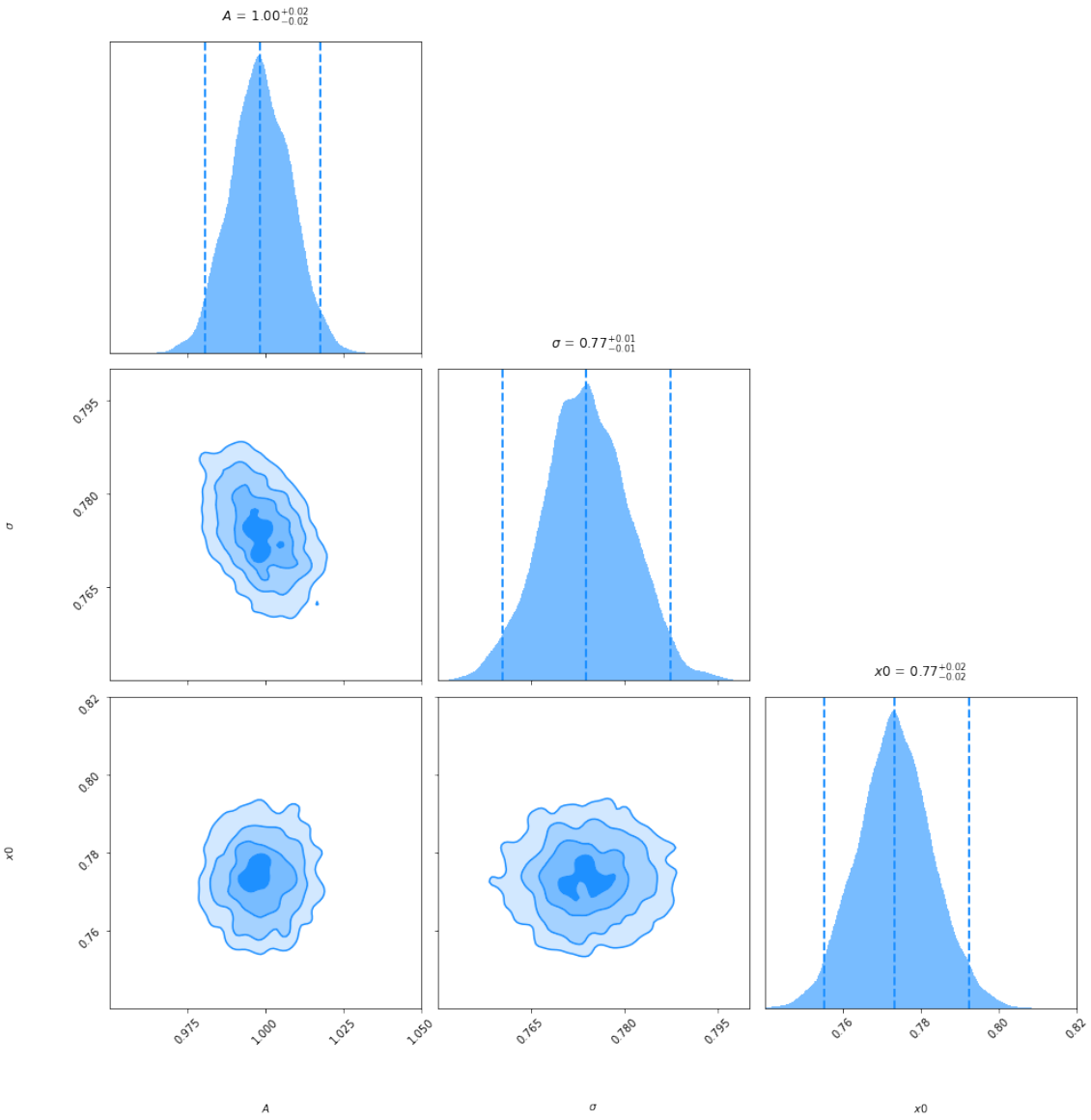


Figure 9. Corner plot from `dynesty` for the case of a spectrum with two underlying emission-line components modeled with a single Gaussian.

C. NGC 2207 CLASSIFICATION PROBABILITY MAP

As mentioned, we obtain a probability of the classification being correct along with the classification. Since this is a binary problem, the probabilities are bound between 0.5 and 1.0. As we see in the figure below (11), the regions of lower surface brightness and spiral arms are classified with a probability generally between 0.85 and 0.9; on the contrary, the bulges are classified with 100% probability. A classification probability of 100% is unrealistic; however, we demonstrated in § 4.3 that spectra more complicated than having a simple single or double emission component are classified as having two components. This likely indicates that the bulge regions are extremely complicated and may need more than two emission components to accurately model. Although a detailed analysis of this is beyond the scope of the paper, the collaboration is studying the bulges of galaxies in the process of a merger in more detail. Importantly, the probabilities vary considerably in regions with low signal-to-noise ratios. Therefore, we suggest masking out these areas when using the network's results for subsequent spectral analysis.

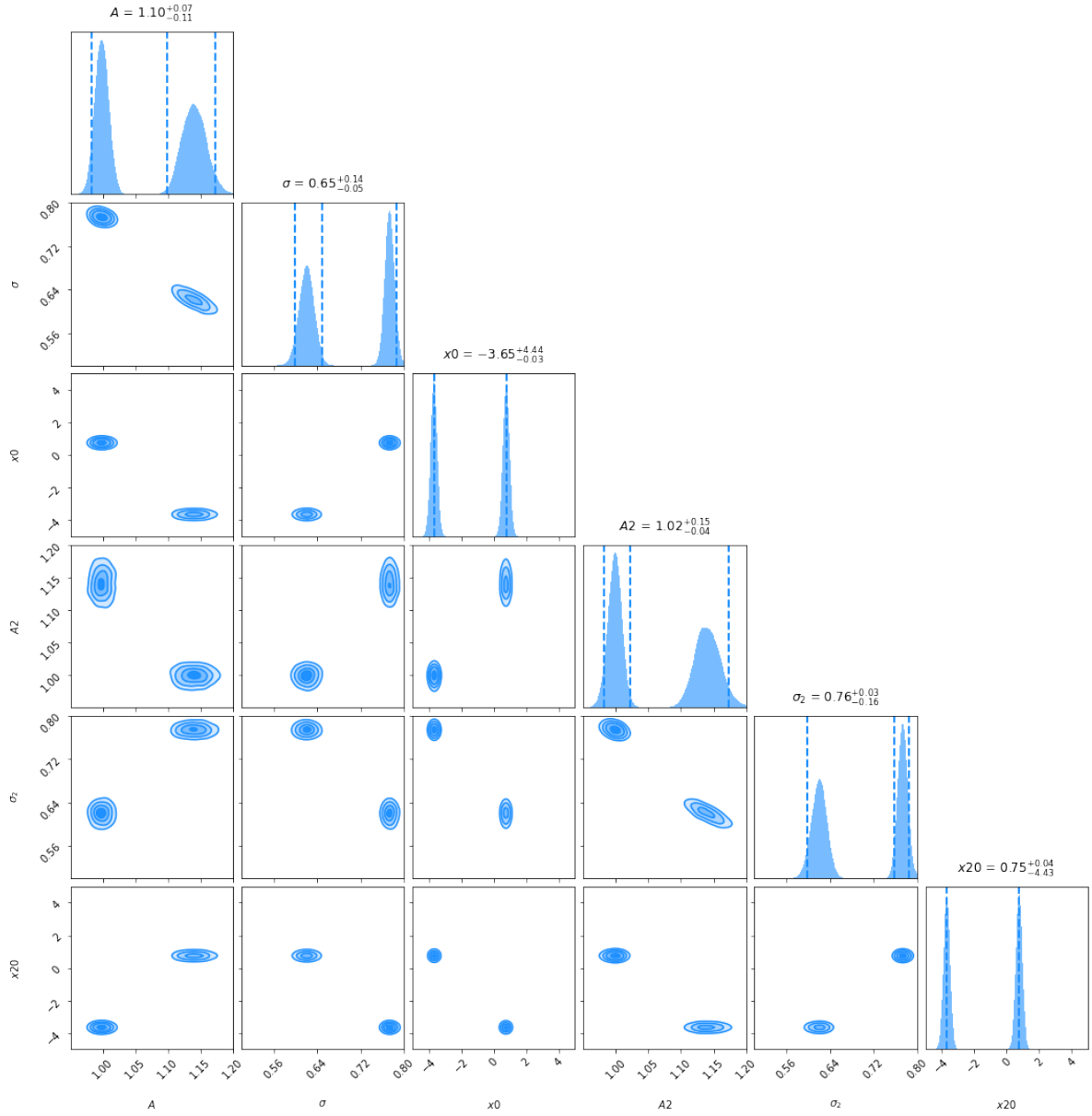


Figure 10. Corner plot from `dynesty` for the case of a spectrum with two underlying emission-line components modeled with two Gaussians.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., et al. 2015, 19

Aniyan, A. K., & Thorat, K. 2017, 230, 20, doi: 10.3847/1538-4365/aa7333

Asari, N. V., Stasińska, G., Morisset, C., & Fernandes, R. C. 2016, 460, 1739, doi: 10.1093/mnras/stw971

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, 93, 5, doi: 10.1086/130766

Baril, M., Grandmont, F., Mandar, J., et al. 2016, 9908

Barnes, J. E. 2002, 333, 481, doi: 10.1046/j.1365-8711.2002.05335.x

Barnes, J. E., & Hernquist, L. 1996, 471, 115, doi: 10.1086/177957

Baron, D. 2019

Breiman, L. 2001, 45, 5

Chollet, F. 2015. https://keras.io

Chopin, N., & Robert, C. P. 2010, 97, 741, doi: 10.1093/biomet/asq021

Conselice, C. J. 2014, 52, 291, doi: 10.1146/annurev-astro-081913-040037

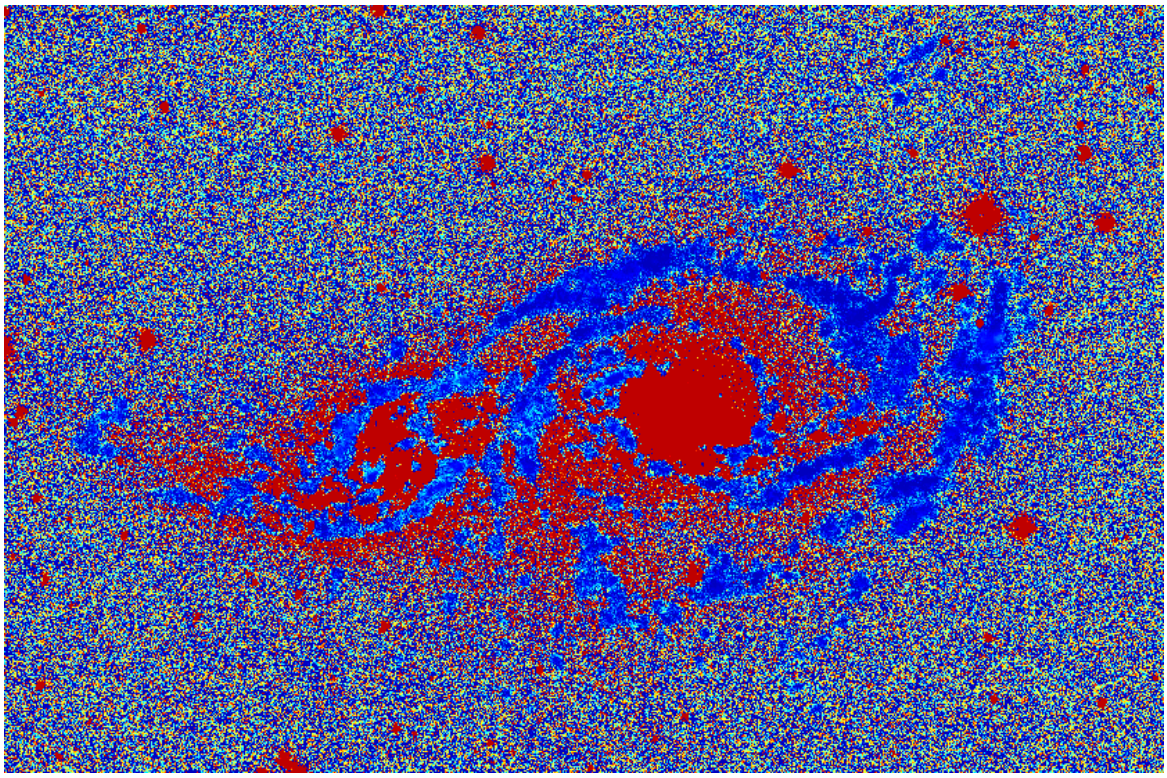


Figure 11. Classification probability map of the NGC 2207 merging system. Red regions correspond to a probability of 1, while regions in dark blue correspond to a probability of approximately 0.85.

Crawford, C. S., Allen, S. W., Ebeling, H., Edge, A. C., & Fabian, A. C. 1999, 306, 857, doi: 10.1046/j.1365-8711.1999.02583.x

Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010, 401, 1552, doi: 10.1111/j.1365-2966.2009.15786.x

Davies, A., Serjeant, S., & Bromley, J. M. 2019, 487, 5263, doi: 10.1093/mnras/stz1288

Donzelli, C. J., & Pastoriza, M. G. 2000, 120, 189, doi: 10.1086/301420

Drissen, L., Rousseau-Nepton, L., Lavoie, S., et al. 2014, 2014, 1, doi: 10.1155/2014/293856

Drissen, L., Martin, T., Rousseau-Nepton, L., et al. 2019, 485, 3930, doi: 10.1093/mnras/stz627

Elmegreen, D. M., Elmegreen, B. G., Kaufman, M., et al. 2017, 841, 43, doi: 10.3847/1538-4357/aa6ba5

Fabbro, S., Venn, K., O'Briain, T., et al. 2018, 475, 2978, doi: 10.1093/mnras/stx3298

Foster, C., Mendel, J. T., Lagos, C. D. P., et al. 2020, 2011, arXiv:2011.13567. http://adsabs.harvard.edu/abs/2020arXiv201113567F

González-Gaitán, S., deSouza, R. S., Krone-Martins, A., et al. 2019-01-21, 482, 3880, doi: 10.1093/mnras/sty2881

Hampton, E. J., Medling, A. M., Groves, B., et al. 2017, 470, 3395, doi: 10.1093/mnras/stx1413

Hunter, J. D. 2007, 9, 90, doi: 10.1109/MCSE.2007.55

Jenkins, C. R., & Peacock, J. A. 2011, 413, 2895, doi: 10.1111/j.1365-2966.2011.18361.x

Kaufman, M., Grupe, D., Elmegreen, B. G., et al. 2012, 144, 156, doi: 10.1088/0004-6256/144/5/156

Keown, J., Di Francesco, J., Teimoorinia, H., Rosolowsky, E., & Chen, M. C.-Y. 2019, ascl:1909.009. http://adsabs.harvard.edu/abs/2019ascl.soft09009K

Kewley, L., Heislerr, C., & Dopita, M. 2001, 132, 37

Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, 57, 511, doi: 10.1146/annurev-astro-081817-051832

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. 2020, 53, 5455, doi: 10.1007/s10462-020-09825-6

Kieseppä, I. A. 1997, 48, 21, doi: 10.1093/bjps/48.1.21

Kim, E. J., & Brunner, R. J. 2017, 464, 4463, doi: 10.1093/mnras/stw2672

Kreckel, K., Ho, I.-T., Blanc, G. A., et al. 2019, 887, 80, doi: 10.3847/1538-4357/ab5115

—. 2020, 499, 193, doi: 10.1093/mnras/staa2743

Kuo, C.-C. J. 2016

Liu, M., Shi, J., Li, Z., et al. 2016

Martin, T., & Drissen, L. 2017

Martin, T., Drissen, L., & Joncas, G. 2012, in Software and Cyberinfrastructure for Astronomy II, Vol. 8451 (International Society for Optics and Photonics), 84513K, doi: 10.1117/12.925420

Martin, T. B., Prunet, S., & Drissen, L. 2016, 463, 4223, doi: 10.1093/mnras/stw2315

McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference, Vol. 1 (Scipy Publishing), 56–61, doi: 10.25080/Majora-92bf1922-00a

Morisset, C., Delgado-Inglada, G., & Flores-Fajardo, N. 2015, 51, 19

O'Briain, T., Ting, Y.-S., Fabbro, S., et al. 2021, 906, 130, doi: 10.3847/1538-4357/abca96

Osterbrock, D., & Ferland, G. 1989, Astrophysics of gaseous nebulae and active galactic nuclei, 1st edn. (University Science Books)

Pooley, C. M., & Marion, G. 2018, 5, 171519, doi: 10.1098/rsos.171519

Rhea, C., Hlavacek-Larrondo, J., Perreault-Levasseur, L., Gendron-Marsolais, M.-L., & Kraft, R. 2020a, 160, 202, doi: 10.3847/1538-3881/abb468

Rhea, C., Rousseau-Nepton, L., Prunet, S., et al. 2021, 2102, arXiv:2102.06230. http://adsabs.harvard.edu/abs/2021arXiv210206230R

Rhea, C. L., Rousseau-Nepton, L., Prunet, S., Hlavacek-larrondo, J., & Fabbro, S. 2020b, 901

Rich, J. A., Kewley, L. J., & Dopita, M. A. 2015, 221, 28, doi: 10.1088/0067-0049/221/2/28

Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, 558, A33, doi: 10.1051/0004-6361/201322068

Rousseau-Nepton, L., Martin, R. P., Robert, C., et al. 2019, 489, 5530, doi: 10.1093/mnras/stz2455

Ruffio, J.-B., Mawet, D., Czekala, I., et al. 2018, 156, 196, doi: 10.3847/1538-3881/aade95

Sereno, M. 2016, 455, 2149, doi: 10.1093/mnras/stv2374

Sharma, S. 2017, 55, 213, doi: 10.1146/annurev-astro-082214-122339

Shields, G. A. 1990, 28, 525, doi: 10.1146/annurev.aa.28.090190.002521

Skilling, J. 2006, 1, 833, doi: 10.1214/06-BA127

Soto, K. T., Martin, C. L., Prescott, M. K. M., & Armus, L. 2012, 757, 86, doi: 10.1088/0004-637X/757/1/86

Speagle, J. S. 2020, 493, 3132, doi: 10.1093/mnras/staa278

Struck, C. 2007, 237, 317, doi: 10.1017/S1743921307001664

Taaki, J., Kamalabadi, F., & Kemball, A. J. 2020, 159, 283, doi: 10.3847/1538-3881/ab8e38

Trotta, R. 2007, 378, 72, doi: 10.1111/j.1365-2966.2007.11738.x

Trouille, L., Tremonti, C. A., Chen, Y.-M., et al. 2013, 477, 211. http://adsabs.harvard.edu/abs/2013ASPC..477..211T

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, 13, 22, doi: 10.1109/MCSE.2011.37

Van Rossum, G., & Drake, F. L. 2009, Python 3 Reference Manual (Create Soace)

Veilleux, S., & Osterbrock, D. E. 1987, 63, 295

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, 17, 261, doi: 10.1038/s41592-019-0686-2

Waskom, M., Botvinnik, O., O'Kane, D., et al. 2017, doi: 10.5281/zenodo.883859

Wild, V., Walcher, C. J., Johansson, P. H., et al. 2009, 395, 144, doi: 10.1111/j.1365-2966.2009.14537.x