Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **470,** 3395–3416 (2017)
Advance Access publication 2017 June 8

doi:10.1093/mnras/stx1413

Using an artificial neural network to classify multicomponent emission lines with integral field spectroscopy from SAMI and S7

E. J. Hampton,[1]★ A. M. Medling,[1,2]† B. Groves,[1] L. Kewley,[1] M. Dopita,[1] R. Davies,[1,3] I.-T. Ho,[1,4] M. Kaasinen,[1] S. Leslie,[1,5] R. Sharp,[1] S. M. Sweet,[6] A. D. Thomas,[1] J. Allen,[7] J. Bland-Hawthorn,[7] S. Brough,[8] J. J. Bryant,[5,7,8] S. Croom,[5,7] M. Goodwin,[8] A. Green,[8] I. S. Konstantantopoulos,[8,9,10] J. Lawrence,[8] Á. R. López-Sánchez,[8,9,10] N. P. F. Lorente,[8] R. McElroy,[5,7] M. S. Owers,[8,9,10] S. N. Richards[5,7,8] and P. Shastri[11]

[1]*Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia*
[2]*Cahill Center for Astronomy and Astrophysics California Institute of Technology, MS 249-17, Pasadena, CA 91125, USA*
[3]*Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstrasse, D-85748 Garching, Germany*
[4]*Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA*
[5]*ARC Centre of Excellence for All-sky Astrophysics (CAASTRO)*
[6]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Mail H30 PO Box 218, Hawthorn, VIC 3122, Australia*
[7]*Sydney Institute for Astronomy (SIfA), School of Physics, The University of Sydney, NSW 2006, Australia*
[8]*The Australian Astronomical Observatory, PO Box 915, North Ryde, NSW 1670, Australia*
[9]*Envizi Group, Suite 213, National Innovation Centre, Australian Technology Park, 4 Cornwallis Street, Eveleigh, NSW 2015, Australia*
[10]*Department of Physics and Astronomy, Macquarie University, NSW 2109, Australia*
[11]*Indian Institute of Astrophysics, Bengaluru 560034, India*

Accepted 2017 June 6. Received 2017 May 22; in original form 2016 June 15

ABSTRACT

Integral field spectroscopy (IFS) surveys are changing how we study galaxies and are creating vastly more spectroscopic data available than before. The large number of resulting spectra makes visual inspection of emission line fits an infeasible option. Here, we present a demonstration of an artificial neural network (ANN) that determines the number of Gaussian components needed to describe the complex emission line velocity structures observed in galaxies after being fit with LZIFU. We apply our ANN to IFS data for the S7 survey, conducted using the Wide Field Spectrograph on the ANU 2.3 m Telescope, and the SAMI Galaxy Survey, conducted using the SAMI instrument on the 4 m Anglo-Australian Telescope. We use the spectral fitting code LZIFU (Ho et al. 2016a) to fit the emission line spectra of individual spaxels from S7 and SAMI data cubes with 1-, 2- and 3-Gaussian components. We demonstrate that using an ANN is comparable to astronomers performing the same visual inspection task of determining the best number of Gaussian components to describe the physical processes in galaxies. The advantage of our ANN is that it is capable of processing the spectra for thousands of galaxies in minutes, as compared to the years this task would take individual astronomers to complete by visual inspection.

Key words: methods: data analysis – techniques: imaging spectroscopy – techniques: spectroscopic – galaxies: general – galaxies: kinematics and dynamics.

1 INTRODUCTION

Integral field spectroscopy (IFS) is changing our approach to studying galaxy evolution. Surveys such as Calar Alto Legacy Integral Field Area (Sánchez et al. 2012), SAMI (Sydney-AAO Multi-object Integral field spectrograph; Croom et al. 2012), Mapping Nearby Galaxies at Apache Point Observatory (Bundy et al. 2015) and S7 (Siding Spring Southern Seyfert Spectroscopic Snapshot Survey; Dopita et al. 2014) are building data bases of spatially resolved spectra of hundreds to thousands of galaxies in order to explore galaxy evolution as a function of morphology, environment and spectral type. IFS provides a powerful probe into the spatial

★ E-mail: elise.hampton@anu.edu.au
† Hubble Fellow

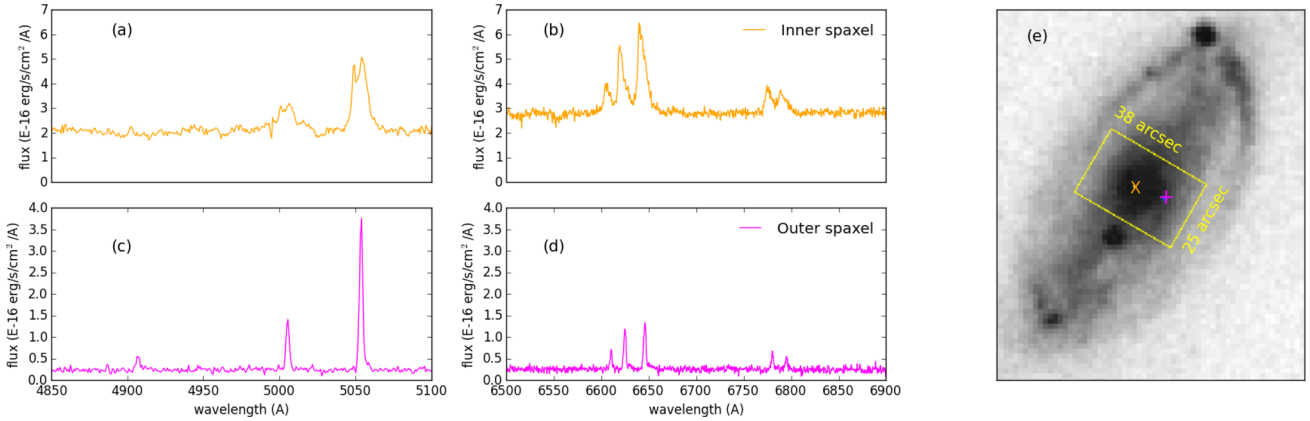

Figure 1. NGC5728. (a) Blue and (b) red spectra from a core spaxel. Note the double peak in the emission lines. (c) Blue and (d) red spectra from a spaxel away from the core of the galaxy. Note the difference in the shapes of the emission lines in comparison to the spectra in (a) and (b). (e) DSS image of NGC5728 with WiFeS field of view overlaid as a yellow box. Magenta '+' indicates the outer spaxel shown in panels (c) and (d), orange 'X' indicates the core spaxel shown in panels (a) and (b).

variation of physical processes across galaxies. For example, single-fibre redshift surveys such as SDSS (Sloan Digital Sky Survey; York et al. 2000) and Galaxy And Mass Assembly (Driver et al. 2011) observe only a single spectrum for each galaxy, typically from the galaxy's core. As such, one can often misidentify the global properties of an individual galaxy (e.g. Fogarty et al. 2012; Ho et al. 2014; Richards et al. 2014, 2016).

Access to the wealth of information from an IFS survey comes at a price: data volume. Not only are advances in IFS technology pushing the previous sample size boundaries but each galaxy observation now contains as many individual spectra as an entire early redshift survey.Data cubes of multiple gigabytes, with thousands of spaxels (spatial pixels) for each galaxy, are not uncommon.

Data reduction pipelines (e.g. Husemann et al. 2013; Sharp et al. 2015; Allen et al. 2015a) are efficient ways to convert raw data into a final spectral data cube for analysis, but interpreting these spectra remains a significant challenge. The data volume is too great to allow tailored analysis of each spectrum individually. Some form of automated analysis is required to extract information from the spectra and to target galaxies for further investigation.

Automated continuum and absorption line fitting is routinely used to understand the stellar populations within galaxies and subsequent emission line fitting provides insight into the star formation, AGN (active galactic nuclei) activity and shock properties of galaxies. This type of pre-analysis can be time consuming for IFS surveys and fitting each emission line by hand is no longer a feasible option. There can be multiple processes behind a single emission line such as emission coming from different regions of a galaxy in the same line-of-sight, multiple kinematic components and/or different excitation mechanisms causing the emission line shapes we observe.

The complexity of emission lines create further steps to our pre-analysis. Automated emission line fitting, including multicomponent fitting for situations with multiple physical processes contributing to emission lines, are currently in use, e.g. LZIFU; Ho et al. (2016a) and GANDALF; Sarzi et al. (2006). Often, a statistic is used to determine where the cut-off is between one or more components, for example χ^2 values or an F-test. However, human input is often still required to decide where to make the cut-off, if one can be made, that separates 1-component fits from higher order fits. When using no statistics, humans are required to inspect each individual spectral fit to decide the best number of components. Manually classifying the spectra of many galaxies is a time-consuming venture but does lead to picking out interesting scientific results that may be missed when using a simple cut-off statistic. This paper describes our study into using an automated machine learning algorithm to remove the time-consuming human input of visually inspecting multicomponent emission line fitting for large surveys by creating an astronomer-like three-way classifier.

2 SPECTRAL PROPERTIES FROM DATA CUBES

Analysis of a galaxy data cube requires the measurement of key physical properties extracted from the individual spectrum of each spaxel. The spectrum at each spaxel typically contains an emission line spectrum, arising from shock-heated or photoionized gas, superimposed upon continuum light, either from the underlying stellar populations or an AGN. Accurate modelling and subtraction of the underlying continuum is critical in correcting for stellar absorption, which would otherwise lead to the incorrect measurement of coincident emission lines, e.g. predominantly those of the Hydrogen Balmer series, H α and H β, in which stellar absorption can have a high equivalent width.

Each line of sight into the galaxy can encompass gas at different velocities and with different excitation mechanisms. For example, Fig. 1 highlights the changing physical processes across a galaxy. Subfigures (a) and (b), (c) and (d) show the variation in emission line profiles between the core (orange) and off nuclear region (magenta) of the S7 galaxy NGC 5728. The nuclear spectra show an obvious double-peaked emission line profile that can be attributed to gas moving in two directions, e.g. an outflow of gas in the disc of the galaxy. Multiple Gaussians are fit to these resulting composite emission lines in order to explain the underlying physical processes occurring within a galaxy. However, we do not a priori know the number of physical components within a resolution element.

To fit the spectral data cubes, we use the automated fitting package LZIFU (Ho et al. 2016a). This program, written in the IDL programming language, fits multiple Gaussian components to each emission line complex in a spectrum after correcting for the underlying stellar absorption component using PPXF (Cappellari & Emsellem 2004). The emission lines are fit simultaneously, forcing each component to have a single velocity and velocity dispersion across all line transitions. An example of the Gaussian fitting is presented in Fig. 3. However, the relative fluxes of the emission lines are left free, and

the line ratios for each component can vary (for a full description see Ho et al. 2016a).

S7 (Dopita et al. 2014) and SAMI (the SAMI Galaxy Survey; Croom et al. 2012; Bryant et al. 2015) are two example galaxy surveys we have used to determine the validity of using machine learning in this application. See Section 4 for details on the S7 and SAMI surveys. For each galaxy observed we have ∼1000 spaxels, each with an associated high spectral resolution spectrum. Each spectrum is fit in turn with 1-, 2- and 3-Gaussian components for the strongest emission lines by LZIFU. The significant challenge is in identifying which set of Gaussian components best describes the spectrum for each spaxel. Visual inspection to make the identifications, which is the common approach, can take up to 1 h for a single galaxy. For small surveys this manual classification is feasible, but with surveys the size of SAMI (∼3000 galaxies) this means ∼125 d continuous work, or multiple years for a single astronomer.

F-tests are used to compare models fitted to a data set and determine how many Gaussians are required to fit emission lines by testing whether the increase in the χ^2 value of the fit justifies the addition of extra parameters. We have also looked at the precision of using *F*-tests, as was done in McElroy et al. (2015), in comparison to using astronomers and our machine learning algorithm, see Section 6.2. However, in complex parameter spaces such as those involving multiple Gaussians, the chi-squared value produced by least-squares minimization algorithms (such as mpfit Markwardt 2009, which LZIFU uses) may represent a local rather than global minimum. This local minimum can be avoided by implementing algorithms like MCMC or Nested Sampling, which probe a greater proportion of the parameter space. Nested sampling has been tested on SAMI galaxies, and because it samples much of the parameter space as it converges upon a fit model comparison comes for free as we can calculate the integrated likelihood for use with Bayesian model comparison. Unfortunately, performing the fitting and model comparison for complex spectra takes over a minute per spaxel.

3 A SUPERVISED ARTIFICIAL NEURAL NETWORK

The machine learning algorithm we have chosen to implement is an artificial neural network (ANN) designed to learn and make classification decisions across an entire survey.[1] Supervised machine learning covers a wide range of distinct classes of artificial intelligence (AI) such as ANNs, support vector machines (SVMs) and Random Forest algorithms that learn without being explicitly programmed. Each have their benefits and weaknesses but all are based on the same underlying principle, they learn from a training set and create models to be used to predict outcomes. We have chosen to use a supervised ANN to build a classification model for multicomponent emission line fitting.

The use of machine learning in astronomy is not a new idea. Contemporary examples include the prediction of solar flares using a SVM (e.g. Bobra & Couvidat 2015), understanding gamma-ray emission from AGNs using Random Forest and ANNs (e.g. Doert & Errando 2014; Hassan et al. 2013), and the classification of galaxy types using image analysis (e.g. Kuminski et al. 2014).

This study is not the first to use machine learning and spectral fitting. Lindner et al. (2015) developed a method of autonomous Gaussian decomposition for the 21 cm SPectral line Observations of Neutral Gas with the EVLA survey (Murray et al. 2015). Their technique was centred on determining the best initial guesses for the Gaussian fitting, which is often the most difficult part of line fitting (see Ho et al. 2016b for a discussion on selecting initial conditions for emission line spectra) and used a combination of computer vision (often an ANN method) and derivative spectra.

Our study, unlike Lindner et al. (2015), centres on what comes after the fitting process. Instead of determining the number of Gaussians beforehand, we make the determination after the emission line fitting has been conducted with 1-, 2- and 3-Gaussian components. This does mean we increase the computational time required for the emission line fitting. However, we also then have multiple options for the numbers of Gaussians already calculated in case a different training of our ANN finds a different answer to the best number of components later on.

The choice of an ANN is based on the architecture of this machine learning algorithm. An ANN is a non-linear algorithm, as will be shown in this section, and is easily scalable to large data sets. The problem is a non-linear problem, each parameter is related to another and thus selections cannot be made without involving all parameters to some degree. The ANN algorithm discussed here is based on matrix multiplication which can scale as $O(n^{2.087})$ for the Solvay Strassen or $O(n^{2.3737})$ for the Coppersmith–Winograd algorithms, if used. For either of these two algorithms an increase in the number of elements does not increase the computation time. However, our data set does not show a difference in running through a small (single galaxy) data set in comparison to increasing the size of the data set to that of the entire SAMI Survey when calculated in Octave. Thus, an ANN was chosen for this prototype study into selecting the number of Gaussian components.

An ANN has three types of layers; an input layer, hidden layers and an output layer. Each node in the input layer, represented as x_j in the following equation, is a parameter value making up a feature vector.

From the input layer these two parameters are sent into the next layer, the first hidden layer of our ANN. The parameter values are put into a sigmoid function with different weights on each parameter. For each node in the first hidden layer the node performs the calculation described in equation (1) where θ_{ij} are the weights for the node i on the input parameters x_j between the input and first hidden layer.

$$a_i = \frac{1}{1 + \exp\left(\sum_{j=1}^{2} \theta_{ij}^1 x_j\right)}. \quad (1)$$

Each node in the first layer uses the same parameters x_1 and x_2 but different weights θ_{ij}^1 corresponding to the specific node.

Once the values of the sigmoid functions are calculated for each node in the first hidden layer they are passed on to the second hidden layer. The process is repeated using the values calculated from the previous layer and different weights corresponding to the different nodes in the second hidden layer. Equation (2) shows the functional form of the equation calculated in the nodes of the second hidden layer.

$$b_i = \frac{1}{1 + \exp\left(\sum_{j=1}^{3} \theta_{ij}^2 a_j\right)}. \quad (2)$$

At this point we have reached the end of the hidden layers.

The output layer is the layer that determines the classification of our input parameters as classification 1 or classification 2. The

[1] Our ANN, called LZCOMP (Lay-Ze-Components), is in a prototype stage and can be found on Github at https://github.com/EliseHampton/LZComp.

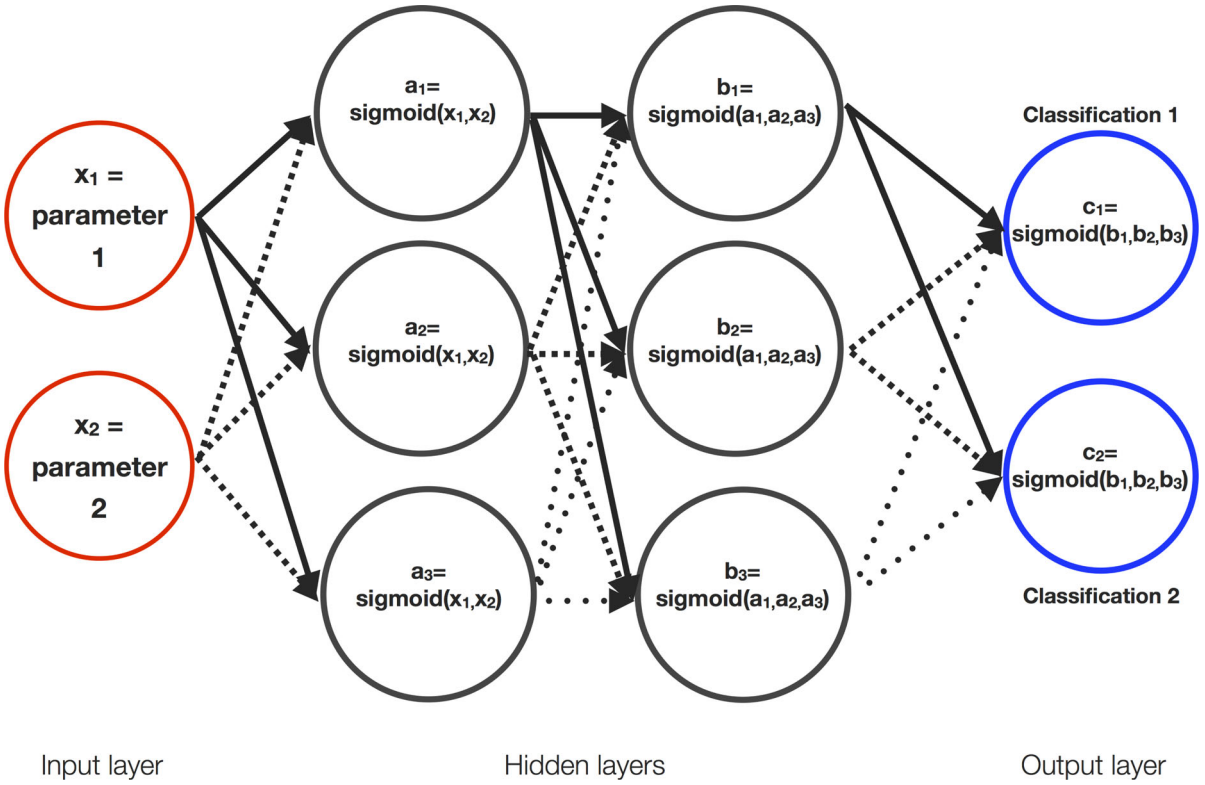

Figure 2. A simple artificial neural network design. The circles indicate the nodes and are grouped into the three layer types: input, hidden and output. This ANN is used to decide if something with the input parameters x_1 and x_2 are of classification 1 or classification 2. At each layer, calculations are done using the values of the previous layer. The final classification is then decided based on which node in the output layer has the largest value.

values b_i from the second hidden layer are sent into the output layer where one last set of sigmoid functions are calculated with a final set of weights. Equation (3) shows the equation calculated by the nodes in the output layer.

$$c_i = \frac{1}{1 + \exp\left(\sum_{j=1}^{3} \theta_{ij}^{3} b_j\right)}. \tag{3}$$

The final classification is determined by which output node has the higher value. During the training phase of the ANN a cost is also calculated. The cost function, equation (4), describes how close the classification from the ANN was to the labels given to the ANN, where m is the number of training examples.

$$J = \frac{1}{m}\sum_{i}(-y_i \log(c_i) - (1 - y_i)\log(1 - c_i)) + \frac{\lambda}{2m}\left(\sum_{ijk}\theta_{ij}^{k}\right). \tag{4}$$

The cost is summed over all output nodes and all training examples. The second term in the cost function sums the weights from each layer with a regularization parameter, λ, that helps prevent any particular weight from becoming too large and dominating the cost function. λ is also known as a tuneable parameter. By changing the value of lambda and comparing the results of the cost function during training we can determine the best value, between 0.01 and 10, to minimize the cost function.

The cost function is minimized in the training phase by iterating over the entire training set using the octave[2] script `fmincg`.[3] The minimization uses the cost function, equation (4), to alter the weights at each node to return a classification closer to the labels for each training example in the next iteration. Each successive iteration adjusts the weights again to create a decision matrix capable of matching the classifications of the labelled training examples.

In the case of our example, in Fig. 2, there are two parameters making up the feature vector but these two values may not be enough to differentiate between the two classifications. For this reason we have to give the ANN enough information to adequately describe the classifications we would like the ANN to make.

Our ANN, LZCOMP, has two hidden layers with 15 nodes in each layer. The input layer has 91 (S7) or 94 (SAMI) input parameters making up the feature vector for each example and the output layer has three nodes corresponding to the best number of components; 1-, 2- or 3-components.

4 TESTING THE ANN ON IFS SURVEYS

Our ANN is a supervised learning algorithm. The supervision comes from training with labelled examples, i.e. we give the ANN the answers for a subset of spectra. The ANN can then use this information to correct itself. The following subsections explain how we have used two test cases, the S7 and SAMI Galaxy Surveys, to test, train

[2] https://www.gnu.org/software/octave/

[3] Originally written by Carl Edward Rasmussen and added to by the Stanford Machine Learning online course. `fmincg` is based on Polack–Ribiere minimization.

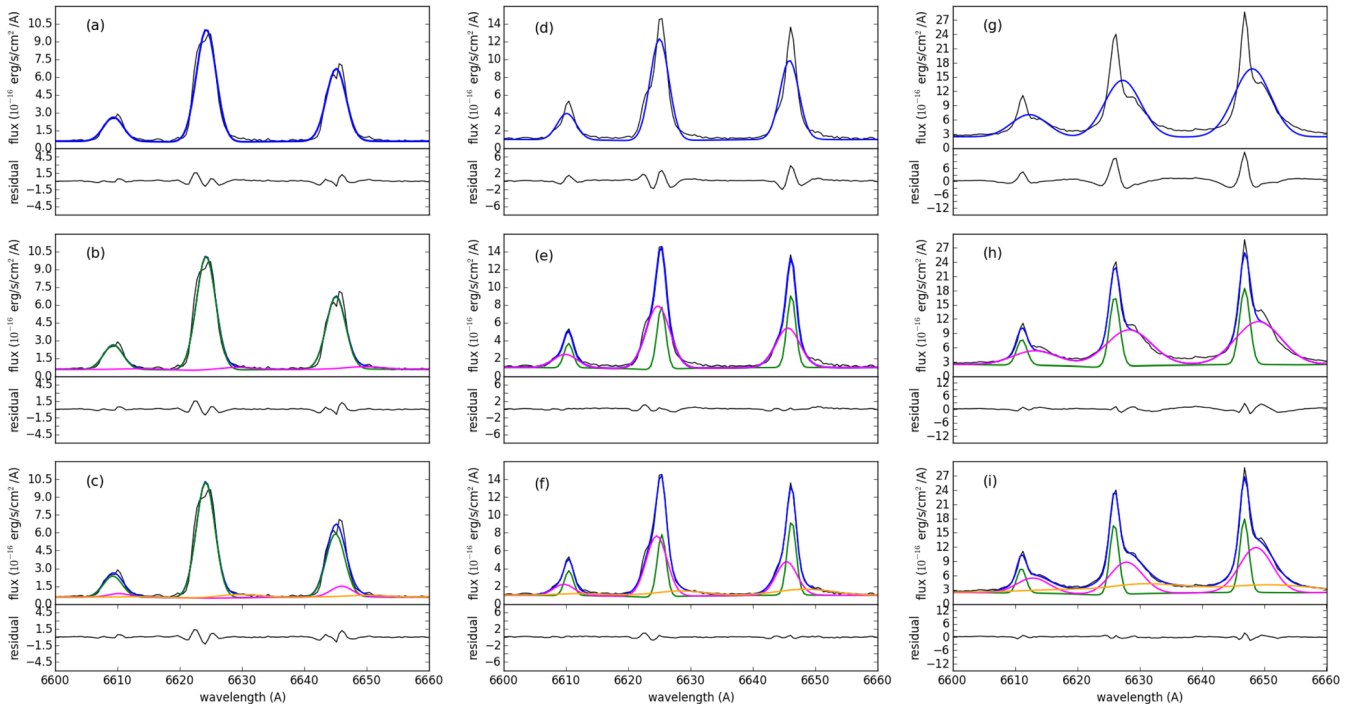

Figure 3. Each column shows an individual spectrum from a different position within the galaxy NGC5728 highlighting the LZIFU fits to [N II]λλ6548, 6584 and H α emission lines where we have high signal-to-noise and high spectral resolution spectra from S7. Each row shows the LZIFU 1-, 2- and 3-component fits to these spectra, respectively. Black shows the data, blue the total line fit, green the 1st component fit, magenta the 2nd component and yellow the 3rd component. The residuals of the total fit are presented below each individual fit plot in black.

and run our ANN in order to classify the number of components needed for each spaxel of a galaxy.

During the testing of the ANN, we found that each survey requires its own training set due to the differences such as the signal-to-noise ratios of the most common emission lines, spectral resolution and the overall galaxy types targeted by each survey. The SAMI Galaxy Survey has been running for over a year and has included studies of galactic winds (Fogarty et al. 2012), studying a unified dynamical scaling relation (Cortese et al. 2014), shocks and outflows within galaxies (Ho et al. 2014), galaxy decomposition (Cecil et al. 2015), star formation and stellar populations (Leslie et al. 2015) and galaxy kinematics (Allen et al. 2015b; Cecil et al. 2016; Cortese et al. 2016; Ho et al. 2016b). S7 has targeted galaxies with very strong AGN-like emission lines to study the physics of the narrow line region in Seyfert galaxies (Dopita et al. 2015; Davies et al. 2016a,b; Scharwächter et al. 2016). When running the ANN on the SAMI Galaxy Survey after being trained with S7, the results showed no correlation with our SAMI trainers, and vice versa.

In both test samples, we have used a range of galaxies to include spaxels dominated by star-formation, known galactic winds, Seyfert 1 and Seyfert 2 AGN-like emission, while making sure we have a range of signal-to-noise ratios that we expect to see in the two samples of galaxies. The input into our ANN during the training phase uses spaxels from eight galaxies randomly put into the training, testing or cross-validation sets of examples (with a ninth unseen galaxy for later testing). All input into our ANN is based on values from LZIFU and are not galaxy specific so we are confident in using spaxels from the same galaxies in both the training and the testing and cross-validation sets, but never the same spaxel. This statement is explained in more detail in Section 6.1.

Using a range of galaxy types with different physical processes to cover the expectations of the entire surveys allows the process of labelling the training data to occur in a short period of time. The expected time to label the eight galaxies is around 8 h for each individual astronomer/trainer. The spaxels are classified as 1-, 2- or 3-components by the groups of astronomers (hereafter 'the Trainers').

In Section 6, we present the number of components selected by the Trainers for both surveys in comparison to how our ANN labels examples, including a comparison to three galaxies unseen in the training phase. In Section 7, we show how using the eight galaxies for training has resulted in physically motivated selections made by our ANN.

4.1 Siding Spring Southern Seyfert Spectroscopic Snapshot Survey : S7

The S7 (Dopita et al. 2015) is a survey of 136 Seyfert galaxies, observed with the Wide Field Spectrograph (WiFeS, Dopita et al. 2010) on the ANU 2.3 m Telescope at Siding Spring Observatory. These galaxies are at redshifts less than 0.02 and thus use most of the field of view of the WiFeS detector (25 arcsec × 38 arcsec) with 1 arcsec pixels. In many S7 galaxies WiFeS only covers the central region of the galaxy. The spectral resolution of S7 is R = 3000 in the blue, and R = 7000 in the red. S7 is intended to explore the narrow and broad line regions in Seyfert galaxies and therefore includes a large number of galaxies with underlying broad emission lines.

Our S7 set of galaxies is made up of 12 galaxies. Eight galaxies are for splitting among the training, testing and cross-validation sets, one galaxy for a check with the results of training and three for a final check and comparison to ensure our ANN is making reasonable selections and the examples are mostly galaxy independent. All of these galaxies are from the initial data release. These galaxies were

chosen to cover the full range of activity within the sample; Seyfert 1's, Seyfert 2's, LINERs and star-forming galaxies.

Manual classification of the training set entails the Trainers (three astronomers from S7) looking at every observed spectrum along with the LZIFU fits using each 1-, 2- and 3-Gaussian components and the resulting residuals. Concentrating on the strong emission lines, the Trainers decides the minimum number of components needed to reproduce the spectrum within the noise. This classification was carried out for every spaxel, resulting in a 2D component mask of the galaxy with values of 1, 2 or 3.

We found that the trainers did not agree for ∼25 per cent of cases. To counteract this disagreement, we trained the ANN using a sample of ∼2900 examples for which at least two of the three astronomers agreed on the number of components. For testing we used another ∼2900 examples split between the cross-validation and testing sets. See Section 5 for an explanation on the numbers of examples used. The choice of using a majority rule is explored in Section 6.1.

4.2 Sydney-AAO Multi-object Integral field: SAMI

The SAMI Galaxy Survey (Croom et al. 2012) is a survey of ∼3400 nearby ($z < 0.1$) galaxies observed with the SAMI instrument on the 3.9-m Anglo-Australian Telescope at Siding Spring Observatory. The survey is made up of four volume-limited galaxy samples with the aim of covering a broad range in stellar mass and environment (Bryant et al. 2015). The survey uses SAMI fibre 'hexabundles' (Bryant et al. 2011, 2014; Bland-Hawthorn et al. 2011) to map these galaxies out to $\gtrsim$1 effective radius. The SAMI hexabundles have 61 fibres and a chosen binning scale of 0.5 arcsec (see Sharp et al. 2015, for details). Each SAMI observation is 15 arcsec in diameter and thus each has fewer spaxels than the S7 galaxies. The spectral resolution of SAMI is R = 4500 in the red and R = 1730 in the blue.

We used 12 galaxies again in our set of galaxies, covering both strong and weak emission line galaxies and Seyfert and star-forming galaxies. Eight galaxies are used in training with a ninth galaxy also labelled for a total galaxy unseen during training and another three galaxies for a final check and comparison as with the S7 extra three galaxies (Section 6). Each spaxel in these 12 galaxies was manually classified by the Trainers (five astronomers from SAMI) in the same manner as with S7. As the galaxies did not always have emission lines towards their outer edges, pixels with no signal were left unclassified.

From the five trainers, we label a spaxel by calculating the most common classification between the five astronomers. This aggregation of labels from all trainers gives us a clean sample of ∼2500 spaxels to train and test our ANN with. We find a ∼50 per cent agreement between trainers. The increased number of trainers corresponds to the lower percentage of agreement among them, compared to S7. The choice of using a majority rule in determining the training and testing labels is explored in Section 6.1.

5 IDENTIFYING THE CORRECT MODEL

The objective of our ANN is to create a reliable, fast, self-consistent and easy to use method of determining the most likely number of components needed to describe a given emission line. As with all supervised machine learning algorithms, we need to train the algorithm before applying it to the survey data. The training involves giving our ANN example input features (vectors containing parameters from the data) and target labels. To create the target labels, we used astronomers to make the decisions for a test set of galaxies (Section 4), labelling each spaxel with the number of components that most likely describe the emission. The ANN is trained with these examples and tested to confirm that the tuneable parameters, e.g. λ, the number of iterations, and the number of training examples obtain the best possible results.

Our approach to training and using our ANN is as follows:

(i) Label a set of galaxies by Trainers for each survey to create labelled examples.
(ii) Create feature vectors, a numerical set of parameters, associated with each example.
(iii) Use half of the eight galaxies' examples to train the ANN to build a model, using the labels to correct the weights. These examples will be the training set.
(iv) Use a half of the remaining labelled examples not used in training as a subset to optimize the tuneable parameters of the ANN. These examples will be the cross-validation set.
(v) Compare results of the ANN to the remaining labelled examples, also unseen in training. These examples are the test set.
(vi) Compare the results of the ANN to each individual trainer using recall and precision.

The test set allows us to calculate the accuracy of our ANN and understand how many more examples we may need in training. The cross-validation set allows us to tune the regularization parameter, λ, to best suit the problem. The testing and cross-validation sets also enables us to understand how many nodes each layer should have to optimize the algorithm and how many layers the ANN needs to give the outcomes that best match our human trainers.

Each example spectrum has 91 (S7) or 94 (SAMI) parameters that describe the emission line fits. These parameters are the parameters related to a fit of multiple Gaussians that we believe to be important for determining if a particular fit is better than another. The parameters include the total signal-to-noise ratio, $\mathrm{Flux_T/Flux_T\_err}$, of each of the strongest emission lines (H α, H β, [N II]λ6583, [S II]λλ6716, 6731 and [O III]λ5007), the relative contribution of each individual component to the total flux of an emission line, e.g. $\mathrm{Flux_1/Flux_T}$. The relative fluxes are important in determining if a particular component is not substantially contributing to the total flux or is contributing a large part of the total flux. A component that contributes little to the total flux may not be significant enough to be a real physical component of the emission line. The input parameters also include the velocity dispersion, σ, in the form of $\sigma/\sigma_{\mathrm{err}}$. By using the ratio of the velocity dispersion with the error in the velocity dispersion we input the certainty of the parameter instead of a galaxy specific value of just σ. Also included are parameters associated with the calculation of the stellar absorption fitting on the underlying stellar continuum; Stellar $\mathrm{EBV^{S7}}$, $\mathrm{H\alpha\_EBV^{SAMI}}$ and $\mathrm{H\beta\_E(B-V)^{SAMI}}$. Our ANN analyses the fitting performed by LZIFU and therefore also includes the goodness of fit parameters as returned by LZIFU; χ^2, χ^2/DOF, χ^2_{cont}. All input parameters that make up the input feature vector used for our ANN are listed in Table 1.

All parameters in our input vector play some role in determining the final result out of the neural network. During the training of the ANN, the weights assigned to each parameter in the nodes determine how much each parameter contributes to the final result. We have not tested which particular parameters, if any, in our feature vectors play the largest roles in determining which label an example results in. However, the weights are designed to make redundant parameters negligible in the final determination of the labels negating the need to go through each parameter and determine if it affects the result during our prototype study of using an ANN.

Table 1. This table presents each input parameter given to our ANN as the input vector. Each flux value is calculated for the following six emission lines: H α, H β, [N II]λ6583, [S II]$\lambda\lambda$6716, 6731 and [O III]λ5007. Also included for each of the six emission lines is the ratio of the total flux from 1-component fitting to the total flux from 2- and 3-component fitting separately. Our ANN is given 91 parameters for the S7 galaxies and is given 94 parameters for SAMI. The difference is in the information given in the fitted cubes before processing through the ANN. The S7 fitting extracts the stellar EBV from the continuum fitting, while SAMI extracts H α_EBV and H β_EBV values from the continuum fit.

	Parameter value included from # component fitting		
Value	1	2	3
$Flux_T/Flux_T$_err	x	x	x
$Flux_1/Flux_1$_err		x	x
$Flux_2/Flux_2$_err		x	x
$Flux_3/Flux_3$_err			x
$Flux_1/Flux_T$	x	x	x
$Flux_2/Flux_T$		x	x
$Flux_3/Flux_T$			x
χ^2	x	x	x
χ^2/d.o.f.	x	x	x
χ^2_{cont}	x	x	x
V_1/V_1_err	x	x	x
V_2/V_2_err		x	x
V_3/V_3_err			x
σ_1/σ_1_err	x	x	x
σ_2/σ_2_err		x	x
σ_3/σ_3_err			x
Stellar EBV^{S7}	x	x	x
H α_EBV^{SAMI}	x	x	x
H β_EBV^{SAMI}	x	x	x

Part of the testing of our ANN has been to identify the best configuration of nodes and layers within the ANN. The number of layers is selected by choosing the minimum number that still successfully minimizes the cost function. Another method to determine how the number of layers changes the results of training is to look at how different layers work alongside each other, while also looking for an adequate number of nodes. Fig. 4 presents the cost function, J (equation (4)), resulting from training using two hidden layers (left-hand side) and three hidden layers (right-hand side) with a different number of nodes in the layers. We have constrained our testing to have the same number of nodes in the two or three hidden layers for each configuration. Each plot shows the results of training using different values of the regularization parameter λ. By comparing the different ANN configurations we can determine which configuration minimizes the cost function. Fig. 4 presents the average cost functions calculated for the training set of examples (blue), the testing set of examples with a regularization parameter matching the training (green), and the cross-validation (CV) set of examples using a regularization parameter of $\lambda = 0$ (red) for comparison. The values are averaged between three randomized sets of the training, CV and testing examples from the SAMI survey. The results from S7 are very similar. Similarly, once the number of layers and nodes have been determined, we determine the regularization parameter, λ, that minimizes the cost function in both training and testing examples. On top of the comparisons made in Fig. 4, we also took into account the accuracy (see Section 6 for an explanation of accuracy in terms of our ANN) of the ANN in selecting 1-, 2- or 3-components in comparison to the training labels.

As is clear from Fig. 4, we have chosen this configuration because the two hidden layer ANN gave better results than the three layer, independent of the number of nodes in each layer. Using a two hidden layer configuration, we found that, after using three different randomized input training, testing and cross-validation sets, 15 nodes gave the lowest training cost functions with different parameters. Once this configuration was set, we narrowed down the testing to a range of regularization parameters, $\lambda = 0.3, 1, 3$ where the cost function sat at a minimum consistently, to perform a more detailed analysis of the training results for each component.

In addition to the cost functions presented in Fig. 4, we determined the configuration of our ANN also based on the success in training of 1-, 2- and 3-components. Our ANN must distinguish between different components in the testing and cross-validation sets of examples to its best ability. The cost function can easily be taken over by the accuracy the ANN has in selecting 1-components, which dominate the spaxels in the testing sets in each survey (as explained in Section 4). Fig. 5 presents the recall and precision values (described in detail in Section 6) as the accuracy of our ANN in selecting the components in comparison to the Trainers for the SAMI survey. For our ANN we need to maximize the recall and precision for each number of components without losing the ability of the ANN to select a specific number of components. For example, a $\lambda = 0.3$ has a high precision for 3-components, but it is lower in precision in selecting both 1- and 2-components that a $\lambda = 3$. The red line presented in Fig. 5 presents the precision and recall of using a regularization parameter of $\lambda = 3$, the value that allows the ANN to best make its selections.

By combining the results in Fig. 4 with the success of the ANN for a two hidden layer, 15 node configuration with a regularization parameter of $\lambda = 3.0$ for our SAMI sample (Fig. 5) and $\lambda = 3.0$ for our S7 sample, we have a configuration that results in an ANN that can select the number of Gaussian components as well as the Trainers.

Not only is the configuration of the ANN important, but also the number of examples used in training. In the case of our ANN, we investigated how many examples would be required to make sure it did not under fit the parameter space leaving us with a model that did not adequately make the classification differences, or overfit the parameter space, leaving a model that can only fit the training examples or does not have the variation in its selections that we see from human selections. Fig. 6 presents the comparison of the cost function of the training set (blue) and cross-validation set (green) as a function of the number of examples used to train. The training cost function is calculated using the regularization parameter assigned to the survey undergoing testing, while the cross-validation cost function is calculated with a regularization parameter $\lambda = 0$. There are two scenarios for determining the number of examples. A high variance system does not converge the two lines, indicating the need for more examples. A high bias system is where the two lines converge to almost the same value, indicating too many examples or the need for more features or hidden layers. The system we are defining needs to be between the high variance and high bias scenarios so as to avoid under or over fitting the training examples. We also need to incorporate enough examples of each 1-, 2- and 3-component fits in order to give the ANN enough examples of each. Fig. 6 presents a system that is in between a high bias and high variance system, showing that the number of examples we are using is adequate for our study. We use $\sim$2400 examples for training for the SAMI Galaxy Survey, and $\sim$2900 examples for S7, to include enough examples of each possible

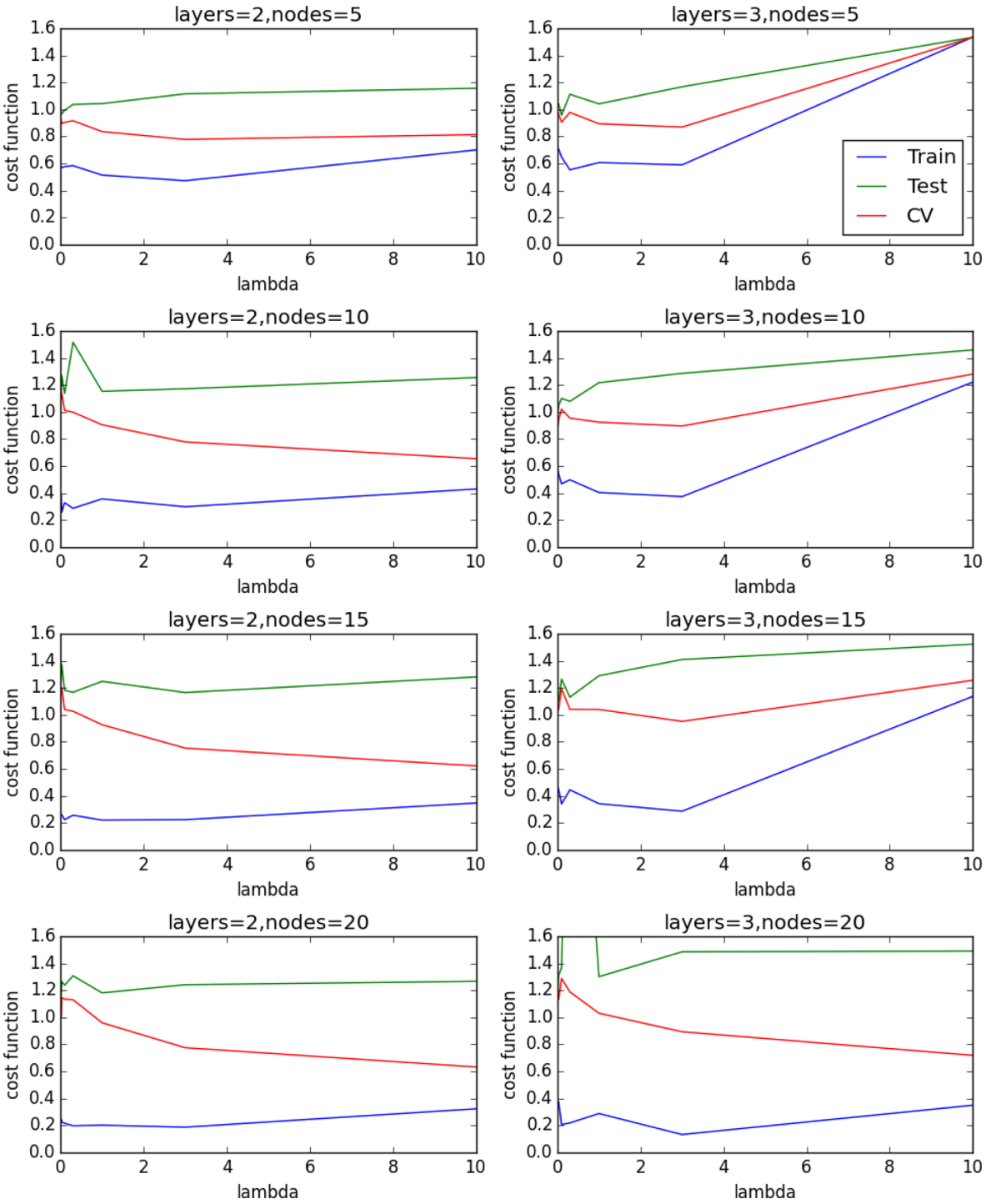

Figure 4. Average cost functions of different configurations of our ANN to determine the best configuration for our study. The results presented here are from the testing of the SAMI Survey. The left-hand side shows the results of a two hidden layer configurations. The right-hand sides presents the results of a three hidden layer configuration. The plots present the cost function of the training, testing and cross-validation sets as a function of the regularization parameter λ of 0.1, 0.3, 1, 3 and 10. The results of the separate randomized examples are presented in Appendix A.

classification and have a model that is between a high bias and high variance model.

The same analysis and testing was completed with the S7 survey with similar results. Section 6 presents the results of accuracy for both survey samples for comparison.

6 ACCURACY

We assessed the accuracy of our ANN after training using two different testing sets. The first testing set, as has been mentioned earlier, is derived from the same galaxies that are used during training, but does not include any spaxels that are used in training, and

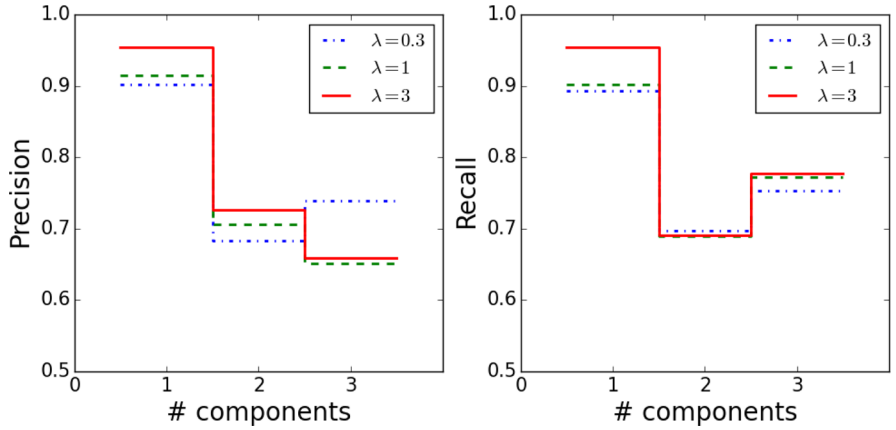

Figure 5. Left: the precision the ANN results with using different values of the regularization parameter λ. Right: the recall the ANN results with using different values of the regularization parameter λ. See Section 6 for a detailed explanation of precision (precision measures how often our ANN will misclassify an example as a particular number of components) and recall (the recall measures the consistency our ANN has for each classification related to how often it misclassifies an example of that component number).

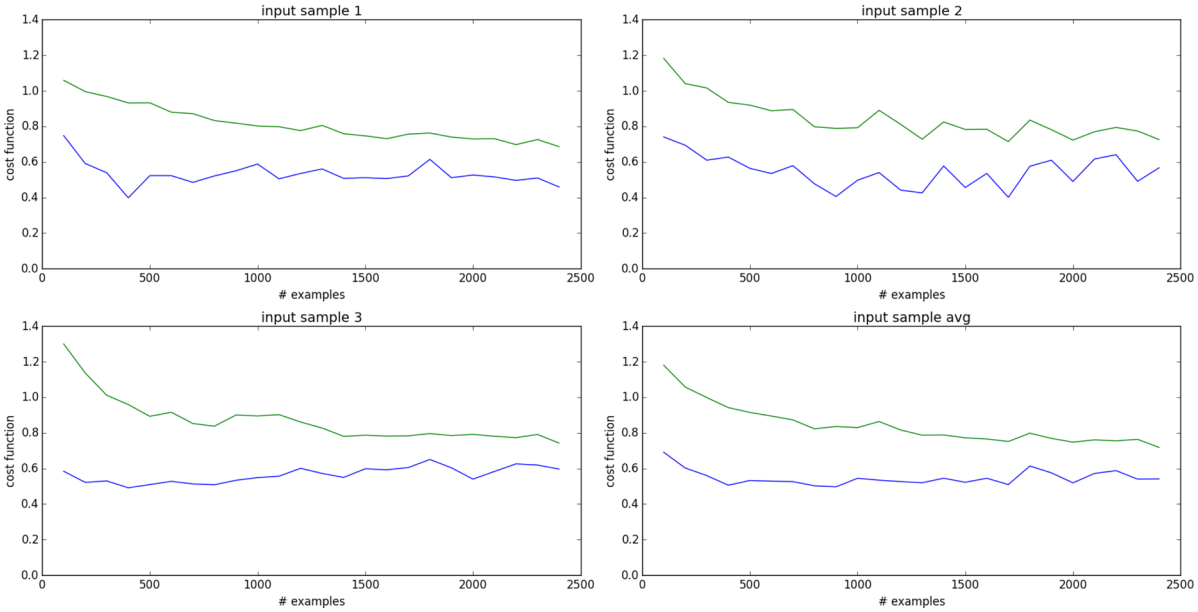

Figure 6. The blue line indicates the cost function of the training set as we increase the number of examples used to train with. The green line indicates the cross-validation cost function. The three input samples are derived from different subsets of the labelled examples assigned to either the training, testing or cross-validation sets. The bottom right plot shows the average cost functions calculated from all three differing sets of examples. The survey testing presented here is the SAMI Survey. Results from the S7 Survey are similar.

one unseen galaxy. The second testing set is three galaxies not seen in training.

The reason behind the two separate testing sets is to show that the spaxels are independent of the galaxies they are derived from as well as a secondary test of how well our ANN makes selections. The first testing set shows similar results to the second testing set, as we show here.

We define the accuracy (how well the ANN can match the labels agreed upon by the Trainers labels) of the ANN as its ability to recall the same classifications as our trainers and the precision in making its decisions.

Equations (5) and (6) show how the recall (R) and precision (P) values are calculated for each number of components. N is the number of examples of which A (our ANN) and T (our trainers) classify with the conditions for A and T as stated in the equations. Together these describe how well using an ANN can classify examples in comparison to our trainers. These values are calculated for each component classification;

$$\mathrm{R_A} = \frac{\mathrm{N_{A=T}}}{\sum \mathrm{N_{A,T=1,2,3}}} \tag{5}$$

$$\mathrm{P_A} = \frac{\mathrm{N_{A=T}}}{\sum \mathrm{N_{A=1,2,3,T}}} \tag{6}$$

More completely, the recall $\mathrm{R_A}$ measures the consistency our ANN has for each classification related to how often it misclassifies an example of that component number. For example, if our ANN correctly classifies 200 examples as 1-components but misclassifies 50 1-component examples as 2- or 3-components, it has a recall of $\mathrm{R_1} = 200/250 = 80$ per cent for 1-component

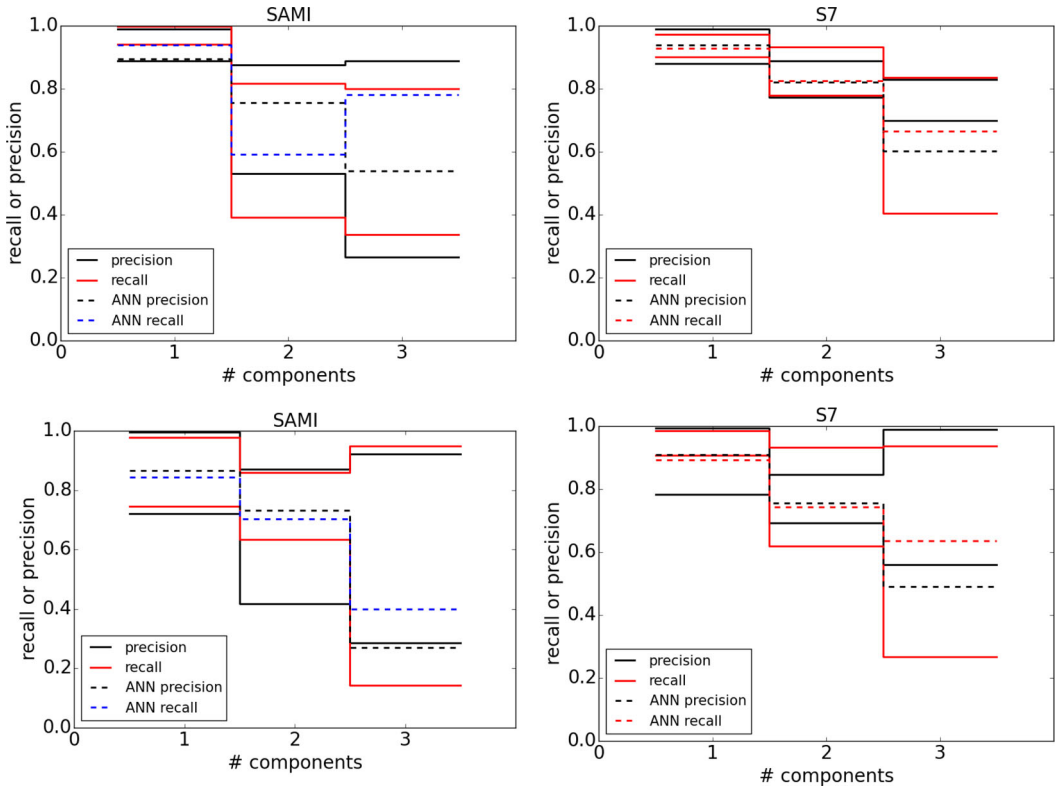

Figure 7. The minimum and maximum values of the precision and recall of our sets of trainers to the precision and recall of our ANN for the SAMI (left) and S7 (right) samples on the first (top) and second (bottom) testing sets. The solid black lines show the min and max values of precision for each component number, the solid red lines indicate the minimum and maximum values of the recall for each component number we have from the trainers in comparison to each other. The dashed lines show results from training using our ANN.

classifications. A recall value is calculated for each classification, 1- to 3-components. The precision P_A measures how often our ANN will misclassify an example as a particular number of components. For example, if our ANN correctly classifies 200 examples as 1-components but also incorrectly classifies 25 examples (of 2- and 3-components) as 1-component, then the ANN has a precision of $P_1 = 200/225 = 89$ per cent for 1-component classifications. Precision values are, like recall values, calculated for each classification.

This same comparison of precision and recall can be repeated using our individual trainers to show how an ANNs performance compares to astronomer's visual inspections. Taking the spaxels from the training galaxies that were not used in training of the ANN and including a completely unseen galaxy, we formed new component maps from $N - 1$ trainers' classifications. We created five new combined classification maps using four of the five trainers for each one successively for SAMI. For S7, we created new combined maps using two of the three trainers successively. The same process as described above is performed on the second testing set for comparison.

In both testing sets we see a large spread in the agreement of classifications through the recall and precision values calculated from our trainers, as is shown in Fig. 7 by the solid lines. We see that the largest spread in the ability of people agreeing with each other is between 2- and 3-components, while agreement is very good over 1-component fits. We also note that our Trainers disagree more often in the second testing set (bottom panels).

The dashed lines in Fig. 7 are the recall and precision values obtained by our ANN. The differences between the top and bottom panels in Fig. 7 are between the resulting precision and recall values. The second testing set uses galaxies not seen during training in any part, while the first testing set does contain spaxels from the same galaxies as used in training (but not the same spaxels). In comparing the precision and recall of LZCOMP we see a drop in the precision and recall values when calculating only over unseen galaxies. However, the agreement between the trainers also changes at the same time.

We designed these tests to show that our ANN is as good as using astronomers. The precision of our ANN in labelling 3-component spaxels does drops below the Trainers minimum precision. However, for S7 the precision is still better than SAMI, thus we agree that this result is suitable. For SAMI the precision of 3-components is not largely different to our Trainers minimum and thus we allow this value of precision.

Overall the results in both cases are mostly within the ranges defined by our trainers. An argument could be made that training on examples in which more trainers agree may increase the precision of LZCOMP, however there are drawbacks which are discussed in Section 6.1.

In addition to the histograms presenting the recall and precision limits of our trainers and ANN, we present the matrices in Fig. 8 to graphically represent the numbers of examples with each possible label. Each element of the matrices contains the total number of examples for that particular label in order to be able to see the numbers of examples in question. The top panels present the matrices for the first testing set of our SAMI (left) and S7 (right) samples. The bottom panels present the matrices for the second testing set (three completely unseen galaxies) for SAMI

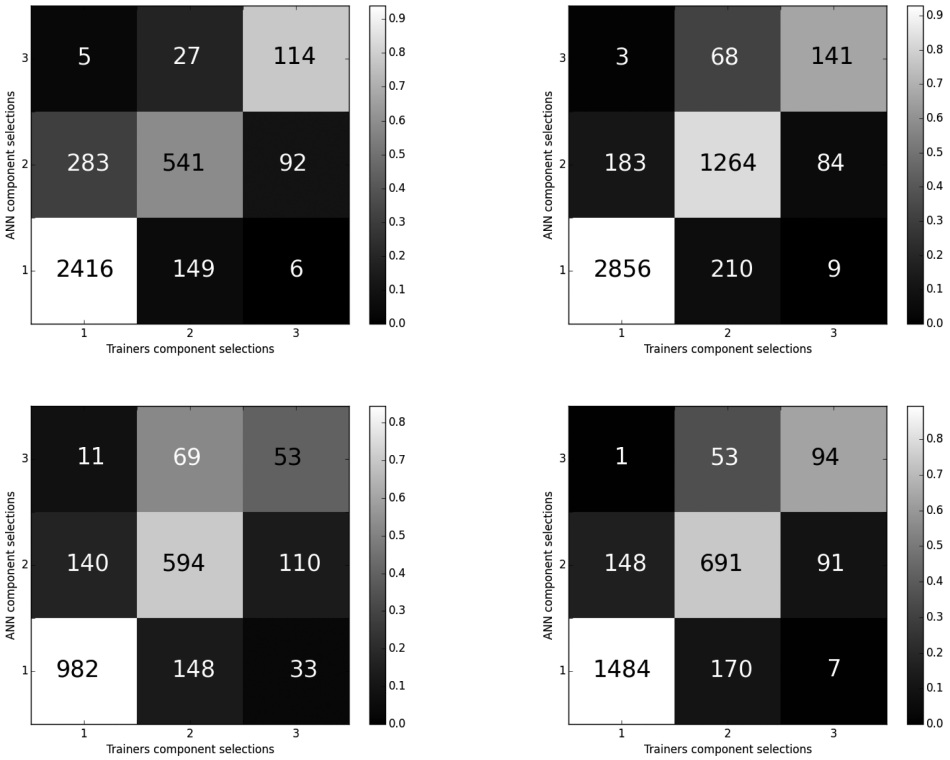

Figure 8. Top: confusion matrices presenting the numbers of examples our ANN selects the same label as the combined trainers for our SAMI (left) and S7 (right) examples. Bottom: the same but for the second testing set of examples. The colours are calculated in the same way as the recall value, measuring the fraction of each example selected by our ANN divided by the total number of the same example labels as chosen by the trainers. An ANN that selects exactly as the combined Trainers would show white along the diagonal and black in the rest. As we found that the Trainers do not agree 100 per cent of the time we do not require our ANN to exactly match the combined Trainers.

(left) and S7 (right). The figures show that our ANN is making the same decisions as the trainers for the majority of examples, even including the second testing set of three unseen galaxies in both surveys.

Using the models learned during training, the ANN is able to successfully classify new spectra. To show this result explicitly, Figs 9 and 10 present the component maps defined by our ANN and the trainers for an S7 galaxy and a SAMI galaxy, respectively. These galaxies were not used for training or cross-validation. We see that the ANN defines a component map that is between all of our trainers maps in both the SAMI and S7 galaxies.

Our ANN's recall and precision is as good as our human trainers. On the full training sets for SAMI and S7, each trainer selected classifications based on what they are seeing. By using the spaxels for which a majority of trainers agreed we avoid any individual favouritism towards a particular number of components from a trainer. Fig. 11 shows the number of spaxels that each trainer classifies as each number of components for the two testing sets of SAMI (left) and S7 (right). Our ANN classifications are also shown to demonstrate favouritism. In both SAMI and S7, we see that the ANN classifies the components in a similar manner to the trainers, following the average favouritism of the trainers as a whole. Our ANN favours 1- and 2-components, but we can see that two of the three trainers also favour 2-components over 1-components. The case may be that the S7 galaxies do have more 2-component spectra than 1-component spectra. S7 is selected to sample very interesting Seyfert galaxies that we expect to require multicomponent fits to express the data.

6.1 Accuracy of ANN using increased agreement between trainers

The labels of the training galaxies are determined using a majority rule. For S7 this means two or three trainers must agree on the number of components required for each training example. For SAMI this means three, four or five trainers must agree.

In order to look at how the accuracy changes with the number of agreements we have re-trained LZCOMP using examples where the agreement is more than the majority. In each case the accuracy, recall and precision, do change for LZCOMP, however the accuracy of our trainers against each other also changes.

The results of using training examples with all three trainers from S7 in agreements are presented in Fig. 12. The precision and recall are presented in the top two panels for the two testing sets (left – half of galaxies not used in training plus one galaxy not seen in training, right – three galaxies not seen in training). The confusion matrices are presented on the bottom row in the same order as above.

Requiring that all three S7 trainers agree results in halving the number of 3-component examples used in training. The precision and recall of LZCOMP trained with the three agreements does result in an increased recall and precision to LZCOMP trained with the majority rule. However, we must take into account that LZCOMP is not being trained or tested on as many examples. The result of LZCOMP obtaining better recall and precision is not indicative of an improved learning when the numbers of examples are lower. Thus, the recall and precision look better but also encompass a smaller number of examples, weighting the results to higher percentages.

The results are similar for the SAMI training and testing sets. However, requiring all five trainers to agree limits the 3-component

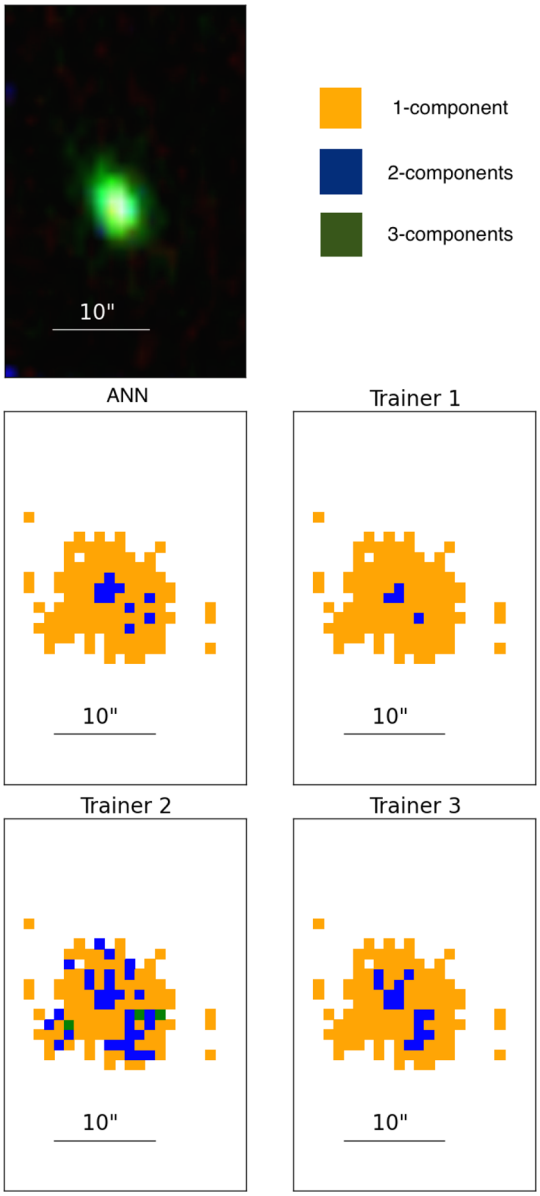

Figure 9. Component maps defined by the ANN and the three S7 trainers for NGC4044 in S7. Green for a 3-component fit, blue for a 2-component fit and yellow for a 1-component fit. Top left image is a 3-colour image created from WiFeS observation; blue [O III], green [N II], red H α emission.

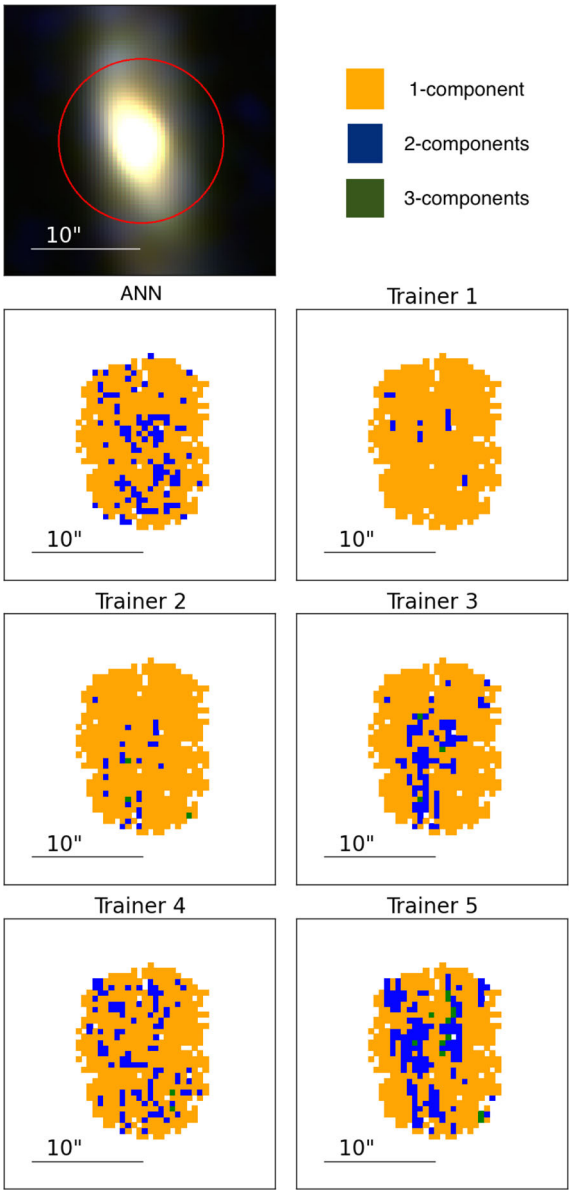

Figure 10. Component maps defined by the ANN and the five SAMI trainers for ID: 9011900367 in the SAMI Galaxy Survey. Green for a 3-component fit, blue for a 2-component fit and yellow for a 1-component fit. Top left image is an SDSS 3-colour image of this galaxy. Red *r* band, green *g* band, blue *u* band. The red circle indicates the SAMI aperture.

training samples to 20 examples. Even though the recall and precision are high they are not representative of LZCOMP's ability to learn. When only requiring four of the five trainers to agree we see the same result as with S7 in that the numbers of 3-component examples is halved. These results are presented in Figs 13 and 14.

Training LZCOMP to select 3-component fits is important. An emission line feature requiring 3-components to describe its shape is indicative of a complex system of gas within that spaxels range. Thus, the choice to allow a majority of trainer agreements means more 3-component examples are used in training, implying a better chance of LZCOMP selecting the 3-component fits on new galaxies.

Although the numbers of 1- and 2-component fits varies up to 50 per cent between the requirements of more agreements, the numbers are not indicative of small number statistics (less than 100). In which case we are confident of the 1- and 2-component learning of LZCOMP in each regime.

6.2 Comparison to the *F*-test

In addition to comparing our ANN results to our trainers, we have also compared the results of using an *F*-test to our trainers. Fig. 15 presents the results of using an *F*-test on our training set of galaxies. The *F*-test selects 1-components more often than our SAMI trainers. The precision obtained by using an *F*-test is comparable to people, but the recall is much lower, pointing to the fact that only ~20 per cent of the spaxels the *F*-test classifies as 1-component agree with our trainers. Due to the discrepancies among our trainers in selecting 2- and 3-components, the *F*-test is comparable in both recall and precision for 2- and 3-components. In comparison to the

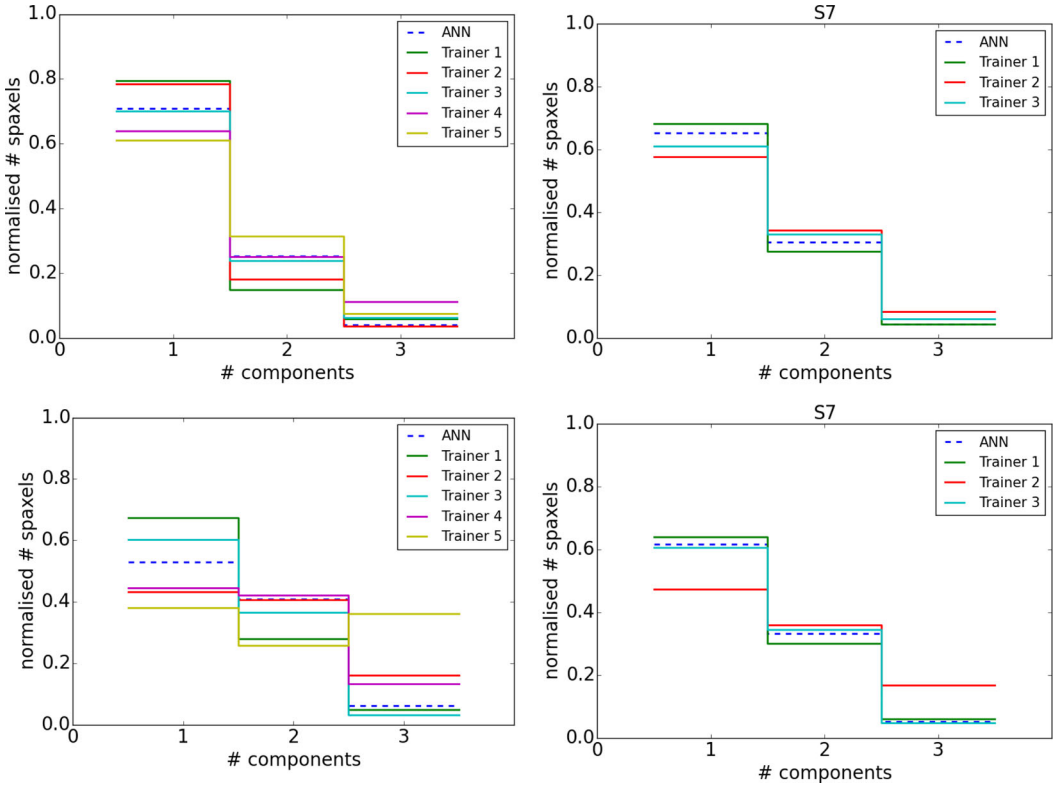

Figure 11. Top: a representation of how many 1-, 2- or 3-component classifications made by each trainer (—) and ANN (–) for the SAMI (left) and S7 (right) on the first training set of galaxies. Bottom: a representation of how many 1-, 2- or 3-component classifications made by each trainer (—) and ANN (–) for the second testing set of galaxies. In both testing sets the ANN predicts similar numbers of 1-, 2- or 3-components for the different surveys.

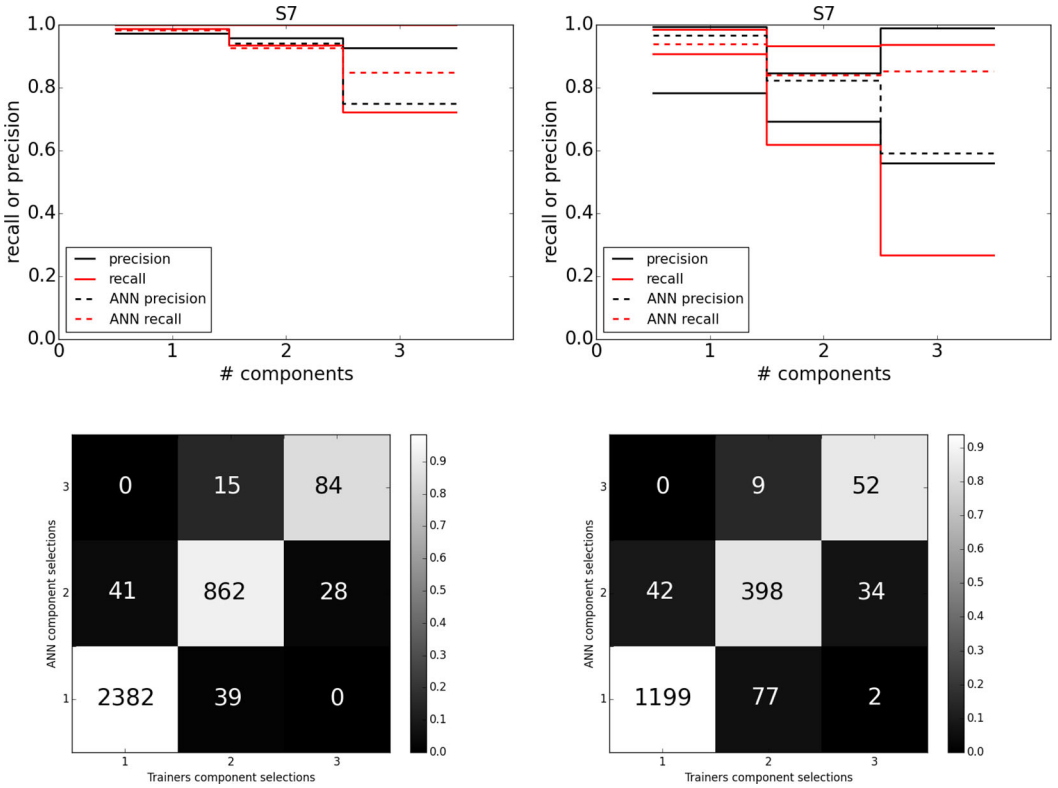

Figure 12. Top: precision and recall values calculated using a training set where all three S7 trainers must agree. Bottom: their relevant confusion matrices. The first testing set is represented on the left and the second testing set is represented on the right.

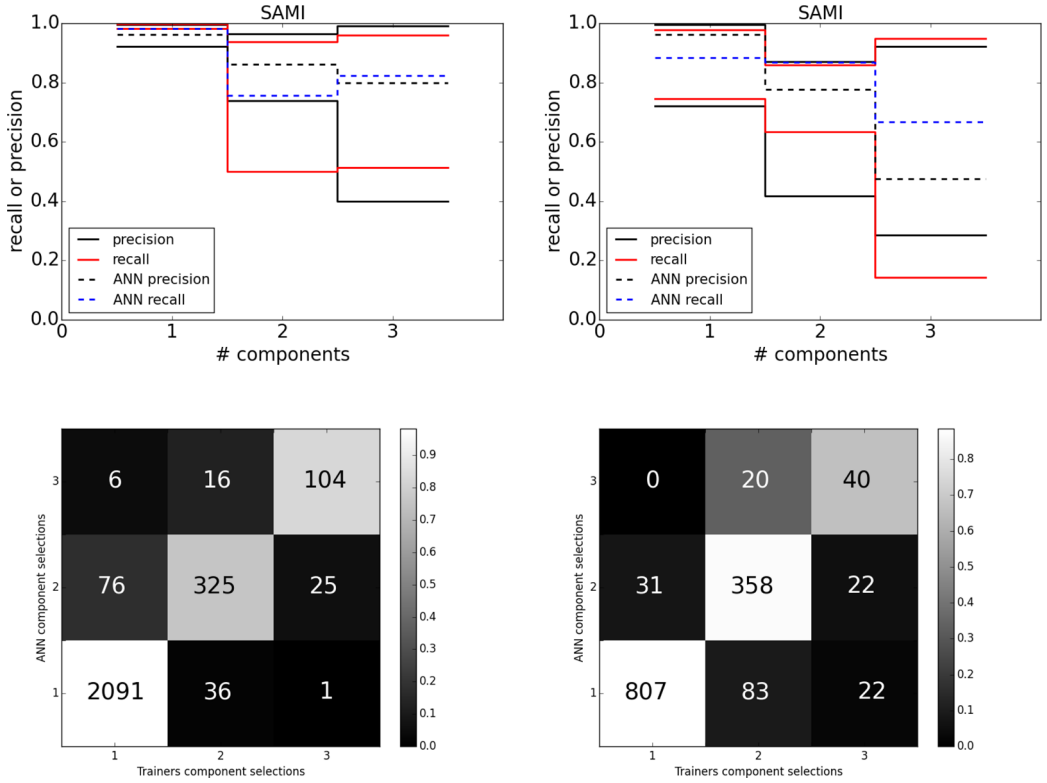

Figure 13. Top: precision and recall values calculated using a training set where at least four of the five SAMI trainers must agree. Bottom: their relevant confusion matrices. The first testing set is represented on the left and the second testing set is represented on the right.

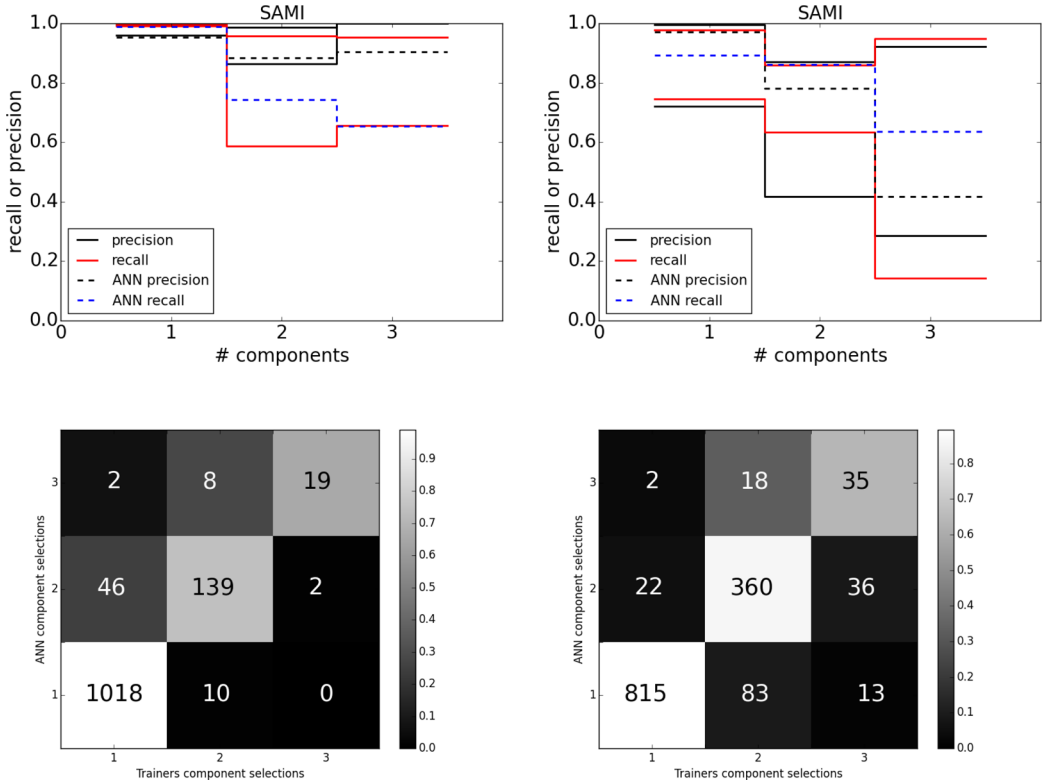

Figure 14. Top: precision and recall values calculated using a training set where all five SAMI trainers must agree. Bottom: their relevant confusion matrices. The first testing set is represented on the left and the second testing set is represented on the right.

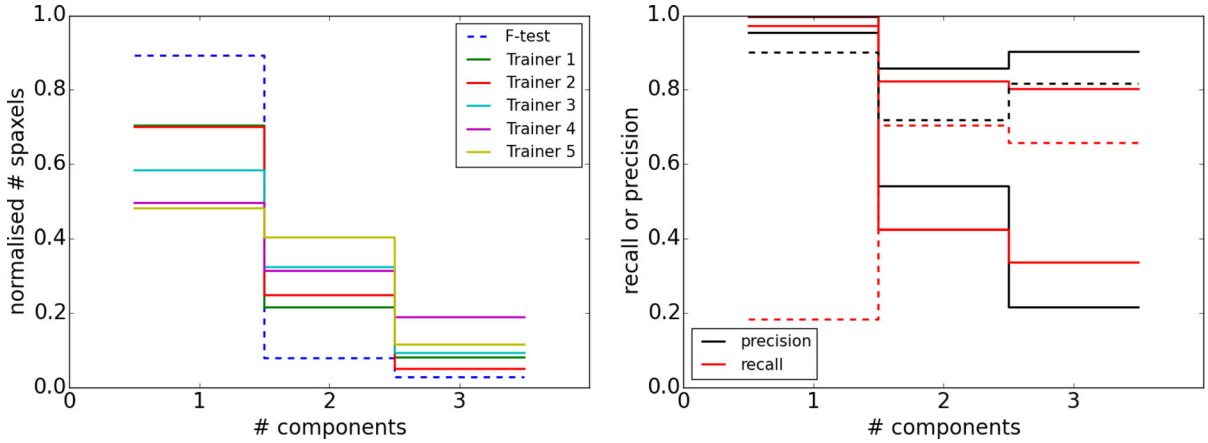

Figure 15. Top: a comparison of the number of 1-, 2- and 3-components classified by the SAMI trainers (—) and using an *F*-test (–). Bottom: a comparison of the recall and precision obtained by using an *F*-test on the SAMI training set of galaxies in comparison to our trainers against each other.

ANN, however, the *F*-test is not as capable at representing another astronomer and their choices of classifications.

7 APPLICATION TO S7 AND SAMI

As presented in the previous sections, the LZIFU code (Ho et al. 2016a), in combination with an ANN, can be used to provide a reliable decomposition of the different emission line components present in galaxies observed with IFS.[4] With the availability of a reliable decomposition analysis for multiple emission lines spanning the full optical spectrum for each of our survey sources, it becomes practical to undertake an in-depth analysis of the wide range of physical processes driving emission with complex composite sources. Early examples of such analysis from the SAMI Galaxy Survey include phenomena such as binary black holes (Allen et al. 2015b), metallicity measurements, corrected for underlying galaxy disc contamination, of isolated H II regions in dwarf galaxies (Richards et al. 2014), and the identification of shocks and outflows in modest luminosity star-forming galaxies (Ho et al. 2014, 2016b).

Fig. 16 shows the *Hubble Space Telescope* (*HST*) image of the Seyfert galaxy NGC7582 (PI: Michael Regan), whose central region has been observed with WiFeS as part of S7 (blue rectangle in Fig. 16). This galaxy has a large star-forming disc, visible in the image. Perpendicular to this disc is an ionization cone with an opening angle of 110 deg that is excited by the central AGN, highlighted in Fig. 17 and described in Dopita et al. (2015). The gas within this cone is highly ionized and extends to 15 kpc. The counter-cone is also visible in our optical observation, but is partly obscured by the dust of the star-forming disc. The red line on Fig. 16 indicates the major and the circle indicates the centre where the AGN is located. The S7 observation of NGC7582 has been fit with 1-, 2- and 3-components using LZIFU and then run through our ANN, to obtain the component maps. The decomposition obtained with LZIFU and our ANN is shown in Fig. 17 for NGC 7582.

The decomposition of emission lines into different components enables the separation of the different excitation processes occurring within a galaxy. In NGC7582, the decomposition of the emission lines separates the galactic disc from the ionization cone and counter-cone. In Fig. 17, panels (a), (b) and (c) present the continuum map, 3-colour emission line total flux map ([O III], [N II] and H α), and the [N II]/H α total flux ratio map of NGC5782. Multicomponent line fitting was necessary because these lines are not always well described by a single Gaussian. The decomposition of NGC7582 is presented in Fig. 17; panels (d), (e) and (f) show the velocities assigned to each component for each spaxel. Component One contains the narrowest emission line components and traces the disc of the galaxy. We can see the rotation curve of this disc gas in panel (g) as we trace the major axis from Fig. 16 along the galaxy. We do not see a turnover in the rotation curve because the S7 observations are looking at only the central regions of the galaxy. The second component, shown in panel (e), consists of the broadest emission line components and traces the ionization cone. We can verify the cause of the broadest emission lines by looking at the velocity of the second component as a function of distance from the centre in the area of the cones. The counter-cone is partly obscured by the galactic disc; we see this in panel (h). The cone and counter-cone are both moving material at a projected velocity of +/− 100 km s^{-1}. The velocity plateaus in panel (h) suggest the front cone is outflowing. We see a possible counter-cone, outlined by blue dashed lines. The remainder of the points in panel (h) are most likely due to the disc of the galaxy broadened due to beam smearing in our line of sight. The third component is a secondary narrow component of emission. In the histograms of velocity dispersions for each component (Fig. 17g), these third components are located between the first and second and are labelled in red. These spaxels have separated narrow peaks with a broader underlying component. These components may be due to the ionization of matter around or at the edge of the cone. To determine what causes this third component, we have looked at the ionization hardness of each component of each spaxel using the [N II] diagnostic diagram (Baldwin, Phillips & Terlevich 1981). Each component is plotted (panel k) in a separate colour, this third component (red) shows high ionization, above the Kewley et al. (2001) diagnostic line. This result suggests that the emission may be shock-induced (Ho et al. 2014; Rich, Kewley & Dopita 2015).

Although this paper does not go into further detail on NGC7582, we have shown that the decomposition of emission lines is important in understanding the detailed kinematics within a galaxy. McElroy et al. (2015) found it beneficial to fit each galaxy with LZIFU then use an *F*-test with harsh cut-offs to determine the component decompositions. In Section 6.2, however, we have shown that the ANN

[4] We note that our study uses the output of using LZIFU, however, there is no reason the same application could not be made to emission line fitting by a code other than LZIFU.

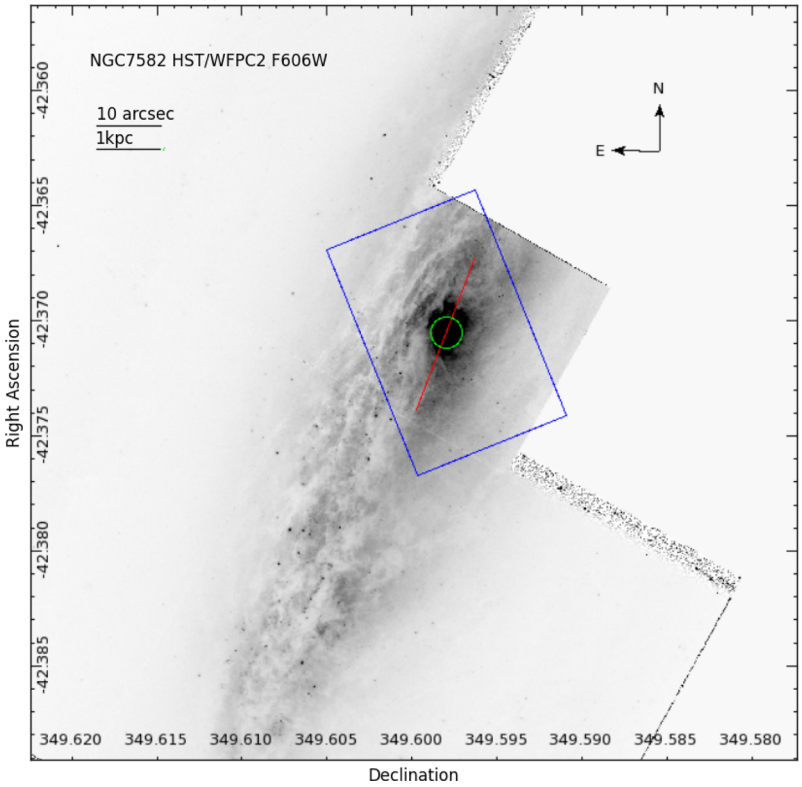

Figure 16. *HST* image of NGC5782. Blue box indicates the S7 FoV and the red line shows the major axis of the galaxy we have used in our analysis of the WiFeS data. The green circle indicates the centre of the galaxy.

is able to more effectively classify complex emission profiles and produces classifications which are more consistent with those of human astronomers.

Surveys are now creating more data than before, meaning that it is not always feasible to fit emission lines by hand, nor to make the component decisions visually. This is where using an ANN is most advantageous. Our ANN is able to process thousands of galaxies and assign the best representation of emission line fits as well as an astronomer in very little time. The quick processing of multicomponent emission line fitting then allows the deeper analysis of galaxies such as NGC7582 through multicomponent emission line fits.

Fig. 18 presents the results of running the SAMI Galaxy Survey DR1 data (Allen et al. 2015c) through our ANN. The top panel shows a histogram of the number of Gaussian components classified by our ANN for all spectra. The total number spaxels fit and classified for SAMI is 348,023, with the majority being one component fits. The bottom panel presents the histograms of fractions of each galaxy that are described by 1-, 2- or 3-Gaussian components. The histogram allows us to pin-point galaxies that show mostly star formation (mostly 1-component fits) and those galaxies that have multiple physical processes ongoing (greater than zero percentage of 2- or 3-component fits).

In a further study, we will be exploring the prevalence of multi-component emission lines in the SAMI Galaxy Survey. This study will consist of comparing the number of emission line components to the galaxy's mass, AGN activity, star formation history and other parameters, to search for correlations that may help in identifying certain types of galaxies or to help understand which types of galaxies contain combinations of certain physical processes. A study of galaxy type and component fitting is only possible with hundreds or thousands of galaxies all fit with multiple components. Using an ANN to make the classifications of the emission line fits has made it possible to do this study on a short time-scale with LZIFU and our ANN, opening the possibility of statistical studies of multicomponent emission processes for a large range of galaxies.

8 CONCLUSIONS

With the larger IFS surveys now in progress, automated complex emission line fitting is a must. LZIFU has automated the fitting process for up to three Gaussian components, but does not have the capacity to determine the best number of components for a particular spaxel. Our ANN provides this capability, showing that the complexities of differentiating between multicomponent fits can be solved reliably and rapidly.

We have built an ANN to take in information produced by LZIFU and output the best-fitting classification for each individual spaxel in each galaxy of a survey. The breakdown of the accuracy into recall and precision of our ANNs results shows that its classifications are indistinguishable from our human trainers. The recall of our ANN for 1-component classifications is 94.2 per cent for S7 and 92.3 per cent for SAMI, while the precision of our ANN for 1-component classifications is 95.3 per cent for S7 and 94.7 per cent for SAMI. The trend of recall and precision for the SAMI test sample continues with 2- and 3-component classifications. For the S7 test sample we see the ANN is capable of matching our trainers in recall but not precision of 3-component classifications. Comparing the precision of the S7 3-component selections to the precision of the SAMI 3-component selections we see the results are very similar between the surveys and so we accept the lower value of 3-component precision in S7 as being adequate. The comparison

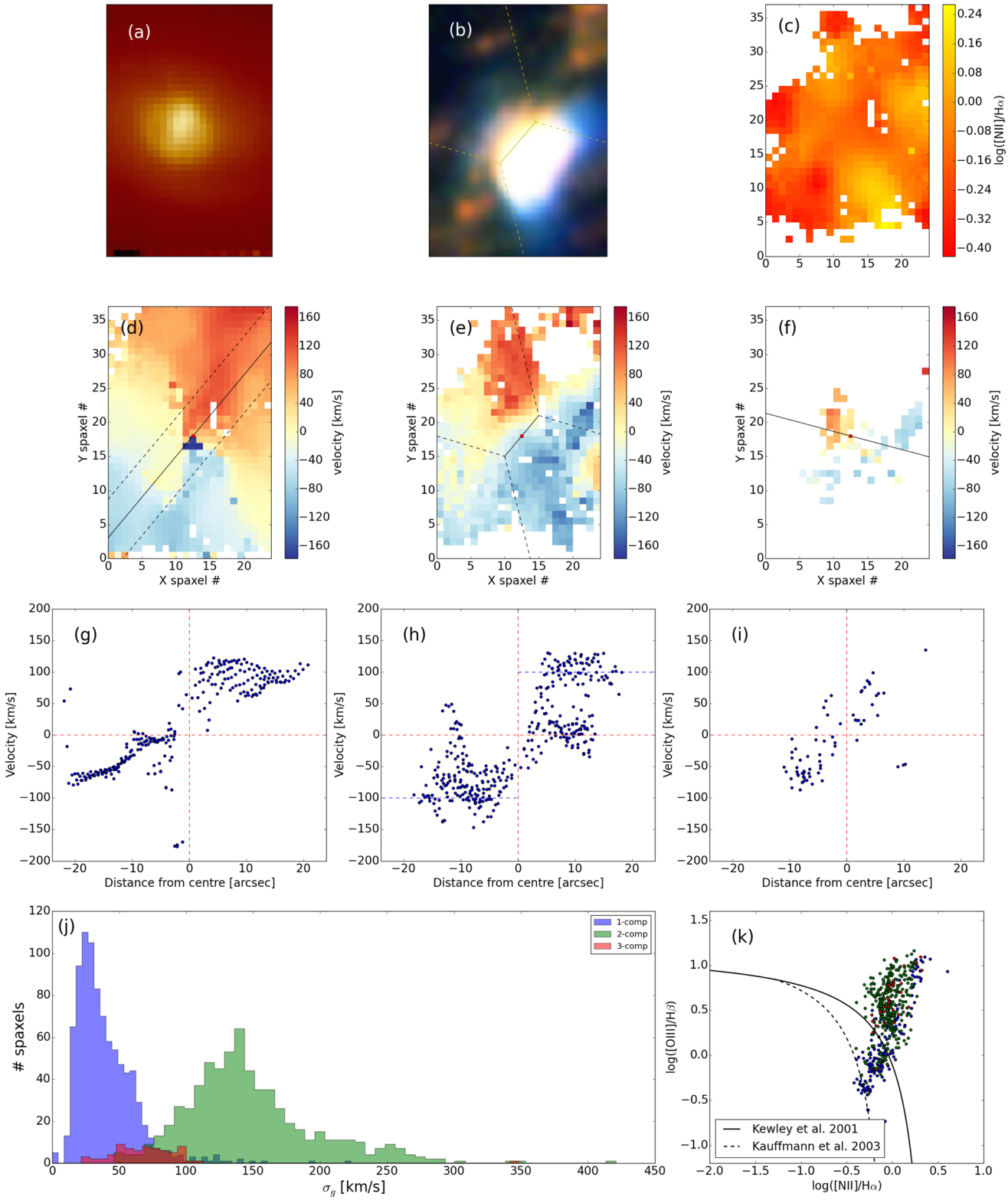

Figure 17. S7 data for Seyfert galaxy NGC7582. (a) Continuum image. (b) 3-colour image of emission lines: [O III] – blue, [N II] – green, H α – red. (c) Total [N II]/H α flux map. (d) Velocity field of the 1st component, related to the disc. (e) Velocity field of the 2nd component, related to the ionization cone and counter-cone. (f) Velocity field of the 3rd component, related to the interaction at the edge of the ionization cone. (g) Rotation curve of the gas disc, points taken from 1st component velocity within dashed lines of (d). (h) Rotation of gas due to the ionization cones. The plateau at +/− 100 km s^{-1} is indicative of outflowing gas. (i) Rotation of 3rd component. (j) Histogram of velocity dispersions of each component. (k) [N II] BPT diagram with components colour-coded to show that the 3rd component has high ionization, as is expected from being caused by the interaction of the ionization cone.

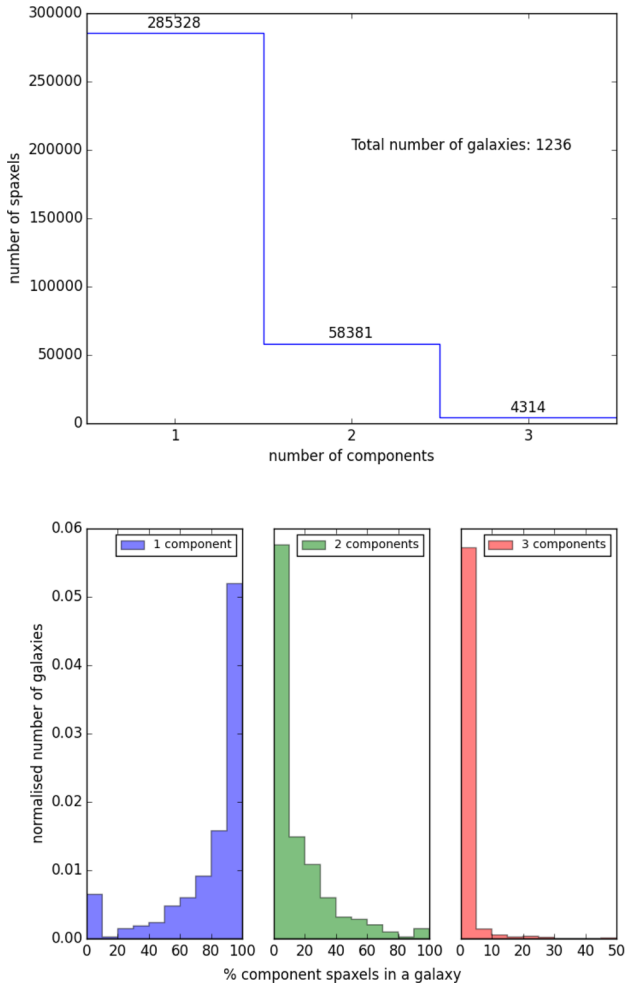

Figure 18. The top histogram presents the number of components classified for the 1236 current SAMI galaxies. The bottom presents the number of galaxies (normalized by the total number of galaxies) with different percentages of 1-, 2- or 3-component spaxels. Although a significant fraction of SAMI galaxies reveal a second component in many of their spaxels, very few galaxies have any spaxels with a third component.

of an ANN with our human trainers, and analysing NGC7582 with the multicomponent fitting results, shows that our ANN approach returns realistic numbers of components.

Our ANN is faster than using astronomers in making the decisions about the best number of Gaussian components need to adequately describe emission lines. For an astronomer, it can take up to an hour per galaxy to make the visual classifications. For small surveys this is a quick process; however, with increased survey sizes the time taken to visually classify every galaxy becomes increasingly long. For example, the time for the final SAMI survey of 3400 galaxies could take 3400 h to visually inspect. In Figs 7, 9 and 10, we showed that astronomers do not agree with each other all the time so using multiple astronomers still does not create a solution.

The time involved in using an ANN greatly decreases the amount of time required to reach the step of scientific analysis. The creation of the training set takes 89 h with multiple astronomers creating labels for the same set of galaxies. The training of the ANN, with up to 50 000 iterations takes 20 min on a 2.7 GHz CPU. Devising of the configuration/regularization parameter is dependant on the programmer but can be up to ∼3 h. To run a survey through our ANN, the majority of the time is taken by reading the information into the ANN and is of the order of minutes for a single 2 GHz CPU. Our study has been to decrease the amount of time required to make the decision on each spaxel of the best number of Gaussian components, which we have done by using an ANN instead of astronomers to make the majority of the decisions.

Our study has used two surveys where the information of each spaxel is put into a text file for easy reading and small file size instead of using a larger file like the original fits files. There is time (of the order of hours) required to put the information from the individual fitting of LZIFU fits files into a final merged fits files with the corresponding information to each number of components selected. For example, the current SAMI Survey of 1236 galaxies takes 16 h on 8 CPUs to be put into their final form. This time to make the final cubes is still required if we were to use visual classification of every galaxy.

The use of an ANN is not limited to classifying the numbers of Gaussian components fit by LZIFU or to the SAMI and S7 surveys. We have shown that an ANN is capable of this classification process and thus could also be used on large IFU surveys to come, where the spectral resolution is high enough to fit multiple components.

ACKNOWLEDGEMENTS

The authors thank the anonymous referee for their helpful comments which have improved the quality of this paper. EJH acknowledges financial support through the Australian National University PhD program and CAASTRO conference support for the ADASS XXV conference 2015. Support for AMM is provided by NASA through Hubble Fellowship grant #HST-HF2-51377 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. BG gratefully acknowledges the support of the Australian Research Council as the recipient of a Future Fellowship (FT140101202). LK and MD acknowledge support from ARC Discovery Project grant (DP16010363). LK acknowledges support from an ARC Laureate Fellowship (FL150100113). SB acknowledges the funding support from the Australian Research Council through a Future Fellowship (FT140101166). SMC acknowledges the support of an Australian Research Council Future Fellowship (FT100100457). MSO acknowledges the funding support from the Australian Research Council through a Future Fellowship (FT140100255).

REFERENCES

Allen J. T. et al., 2015a, MNRAS, 446, 1567
Allen J. T. et al., 2015b, MNRAS, 451, 2780
Allen J. T. et al., 2015c, VizieR Online Data Catalog, 744
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Bland-Hawthorn J. et al., 2011, Opt. Express, 19, 2649
Bobra M. G., Couvidat S., 2015, ApJ, 798, 135
Bryant J. J., O'Byrne J. W., Bland-Hawthorn J., Leon-Saval S. G., 2011, MNRAS, 415, 2173
Bryant J. J., Bland-Hawthorn J., Fogarty L. M. R., Lawrence J. S., Croom S. M., 2014, MNRAS, 438, 869
Bryant J. J. et al., 2015, MNRAS, 447, 2857
Bundy K. et al., 2015, ApJ, 798, 7
Cappellari M., Emsellem E., 2004, PASP, 116, 138
Cecil G. N. et al., 2015, Am. Astron. Soc. Meeting Abstr., 225, 250.10
Cecil G. et al., 2016, MNRAS, 456, 1299
Cortese L. et al., 2014, ApJ, 795, L37
Cortese L. et al., 2016, MNRAS, 463, 170
Croom S. M. et al., 2012, MNRAS, 421, 872

Davies R. L. et al., 2016a, MNRAS, 462, 1616
Davies R. L. et al., 2016b, ApJ, 824, 50
Doert M., Errando M., 2014, ApJ, 782, 41
Dopita M. et al., 2010, APSS, 327, 245
Dopita M. A. et al., 2014, Astron. Astrophys., 566, A41
Dopita M. A. et al., 2015, ApJS, 217, 12
Driver S. P. et al., 2011, MNRAS, 413, 971
Fogarty L. M. R. et al., 2012, ApJ, 761, 169
Hassan T., Mirabal N., Contreras J. L., Oya I., 2013, MNRAS, 428, 220
Ho I.-T. et al., 2014, MNRAS, 444, 3894
Ho I.-T. et al., 2016a, APSS, 361, 280
Ho I.-T. et al., 2016b, MNRAS, 457, 1257
Husemann B. et al., 2013, A&A, 549, A87
Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121
Kuminski E., George J., Wallin J., Shamir L., 2014, PASP, 126, 959
Leslie S., Kewley L., Sadler E., Bryant J., 2015, in Ziegler B. L., Combes F., Dannerbauer H., Verdugo M., eds, Proc. IAU Symp. 309, Galaxies in 3D Across the Universe. Cambridge Univ. Press, Cambridge, p. 21
Lindner R. R. et al., 2015, AJ, 149, 138
Markwardt C. B., 2009, in Bohlender D. A., Durand D., Dowler P., eds, ASP Conf. Ser. Vol. 411, Astronomical Data Analysis Software and Systems XVIII. Astron. Soc. Pac., San Francisco, p. 251
McElroy R., Croom S. M., Pracy M., Sharp R., Ho I.-T., Medling A. M., 2015, MNRAS, 446, 2186
Murray C. E. et al., 2015, ApJ, 804, 89
Rich J. A., Kewley L. J., Dopita M. A., 2015, ApJS, 221, 28
Richards S. N. et al., 2014, MNRAS, 445, 1104
Richards S. N. et al., 2016, MNRAS, 455, 2826
Sánchez S. F. et al., 2012, A&A, 538, A8
Sarzi M. et al., 2006, MNRAS, 366, 1151
Scharwächter J. et al., 2016, in Napolitano N. R., Longo G., Marconi M., Paolillo M., Iodice E., eds, Astrophys. Space Sci. Proc. Vol. 42, The Universe of Digital Sky Surveys. Springer International Publishing, Switzerland, p. 263
Sharp R. et al., 2015, MNRAS, 446, 1551
York D. G. et al., 2000, AJ, 120, 1579

APPENDIX A: DETERMINING THE CORRECT MODEL

Figs A1–A3 present the testing plots used in defining the configuration of our ANN with the three separate randomized input samples.

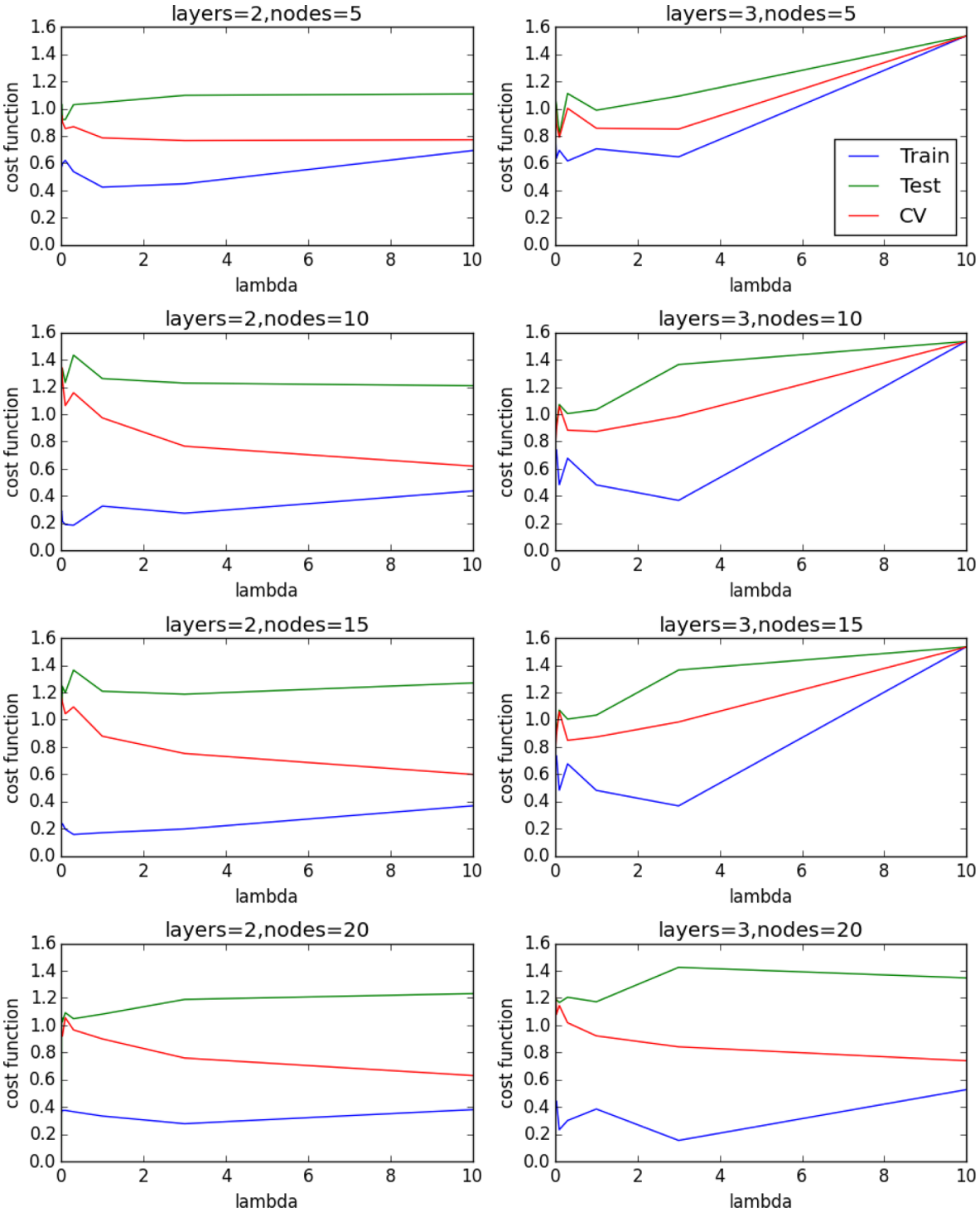

Figure A1. Randomized sample 1: cost functions of different configurations of our ANN to determine the best configuration for our study. The left-hand side shows the results of a two hidden layer configurations. The right-hand sides presents the results of a three hidden layer configuration. The plots present the cost function of the training, testing and cross-validation sets as a function of the regularization parameter λ.

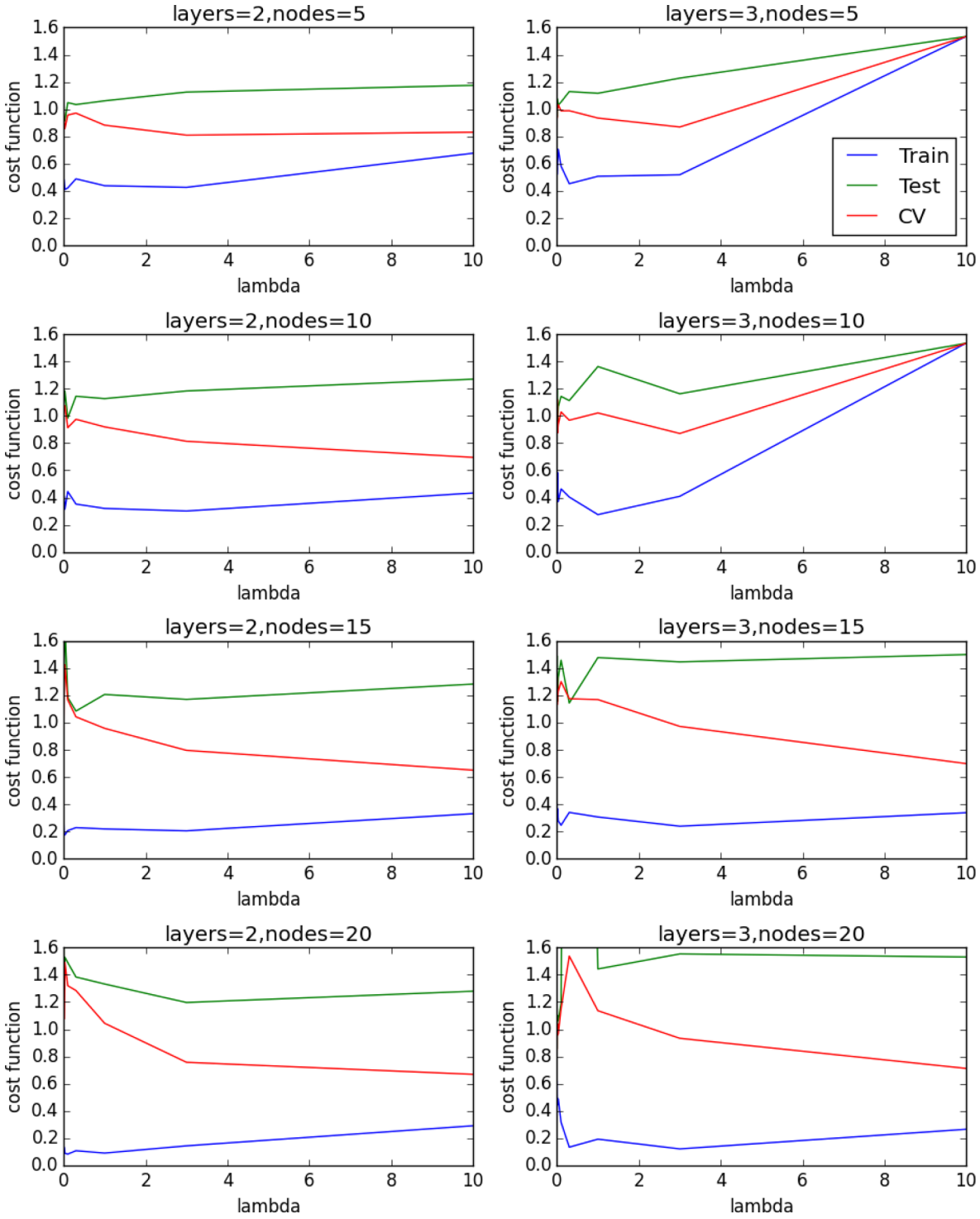

Figure A2. Randomized sample 2: cost functions of different configurations of our ANN to determine the best configuration for our study. The left-hand side shows the results of a two hidden layer configurations. The right-hand sides presents the results of a three hidden layer configuration. The plots present the cost function of the training, testing and cross-validation sets as a function of the regularization parameter λ.

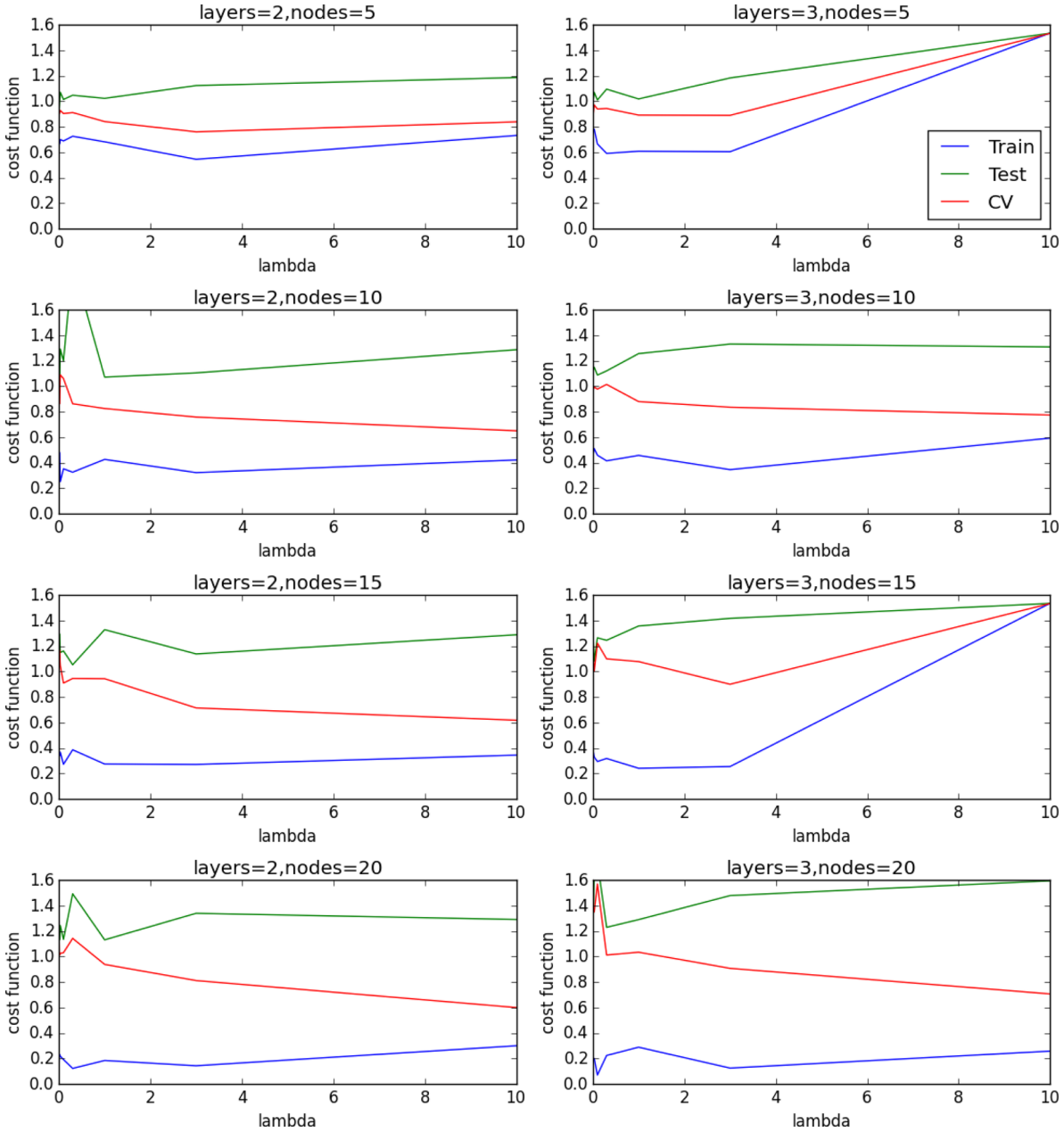

Figure A3. Randomized sample 3: cost functions of different configurations of our ANN to determine the best configuration for our study. The left-hand side shows the results of a two hidden layer configurations. The right-hand sides presents the results of a three hidden layer configuration. The plots present the cost function of the training, testing and cross-validation sets as a function of the regularization parameter λ.

This paper has been typeset from a TEX/LATEX file prepared by the author.