Predicting spectral features in galaxy spectra from broad-band photometry

F. B. Abdalla,1⋆ A. Mateus,2 W. A. Santos,3 L. Sodrè Jr,3 I. Ferreras4 and O. Lahav1

1Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT
2Instituto de Ciencias del Espacio (IEEC-CSIC), Barcelona, Spain
3Departamento de Astronomia, IAG-USP, Rua do Matão 1226, 05508-090, Sao Paulo, Brazil
4Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking RH5 6NT

Accepted 2007 December 14. Received 2007 November 14

ABSTRACT
We explore the prospects of predicting emission-line features present in galaxy spectra given broad-band photometry alone. There is a general consent that colours, and spectral features, most notably the 4000 Å break, can predict many properties of galaxies, including star formation rates and hence they could infer some of the line properties. We argue that these techniques have great prospects in helping us understand line emission in extragalactic objects and might speed up future galaxy redshift surveys if they are to target emission-line objects only. We use two independent methods, Artificial Neural Networks (based on the ANNz code) and Locally Weighted Regression (LWR), to retrieve correlations present in the colour N-dimensional space and to predict the equivalent widths present in the corresponding spectra. We also investigate how well it is possible to separate galaxies with and without lines from broad-band photometry only. We find, unsurprisingly, that recombination lines can be well predicted by galaxy colours. However, among collisional lines some can and some cannot be predicted well from galaxy colours alone, without any further redshift information. We also use our techniques to estimate how much information contained in spectral diagnostic diagrams can be recovered from broad-band photometry alone. We find that it is possible to classify active galactic nuclei and star formation objects relatively well using colours only. We suggest that this technique could be used to considerably improve redshift surveys such as the upcoming Fibre Multi Object Spectrograph (FMOS) survey and the planned Wide Field Multi Object Spectrograph (WFMOS) survey.

Key words: galaxy: formation.

1 INTRODUCTION

Current galaxy formation models can be very successful in predicting the general spectral energy distribution (SED) of galaxies at moderate redshifts by evolving stellar population models and co-adding a certain number of stellar populations with different ages (e.g. Le Borgne et al. 2004). These models have become mainstream in obtaining stellar population properties (like ages) given current data. With them, constraints can be imposed on the formation times of old elliptical galaxies, the epoch of reionization and cosmology (e.g. Jimenez & Loeb 2002; Ferreras & Yi 2004).

However, most stellar population models do not include the modelling of emission lines in galaxies. This is mainly due to the need for modelling the different phases present in the interstellar medium. Some spectrophotometric codes such as Pegase (Le Borgne et al. 2004) do include simple prescriptions for emission lines. Other models such as Starburst99 (Leitherer et al. 1999) actually provide equivalent width (EW) and flux estimates for the most important emission lines usually observed in galaxy spectra.

The broad-band shape of the SED distribution can be related empirically to the presence of an emission line. Roughly speaking, we know that the emission-line strength depends on the condition of the gas inside a galaxy; this is strongly related to the amount of star formation ongoing in this galaxy; the star formation can be inferred to a certain extent from the colour, i.e, we know that the SED of a star burst galaxy is much flatter than the corresponding SED of a more passively evolving red galaxy. Therefore, the shape of the SED must be correlated in some way to the presence of a strong emission feature in the same galaxy.

We propose to find empirical correlations between the EWs of different lines and the broad-band colours from imaging data by utilizing novel statistical methods. One method, Artificial Neural Networks, is based on the ANNz code (e.g. Collister & Lahav 2004; Abdalla et al. 2007; Boris et al. 2007), previously used to predict photometric redshifts from galaxy colours. The second statistical method we explore here is Locally Weighted Regression (LWR).

⋆E-mail: fba@star.ucl.ac.uk
We can then use such correlations to better understand the process of galaxy formation and the conditions under which strong lines are produced. We suggest that with these statistical results, any model attempting to predict emission features in galaxy spectra could be considerably improved by being calibrated to have the same statistical properties as found in nature.

Another strong motivation for this work is the predictability of an emission feature present in high-redshift galaxies. The Sloan Digital Sky Survey (SDSS) and the 2dF Galaxy Redshift Survey (2dFGRS) have probed the nearby Universe. More current [VIMOS-VLT Deep Survey (VVDS) and Deep Extragalactic Evolutionary Probe 2 (DEEP2)] as well as future efforts will attempt to probe a large part of the high-redshift Universe. More specifically, if one is only interested in redshift surveys and not in the more specific SED of a galaxy, then it is important to minimize the amount of time spent looking for a galaxy redshift. The most efficient way of doing this is to perform redshift surveys targeted at emission-line objects. If there is a possibility of statistically predicting which galaxies have strongest emission lines, then the potential time spent locating all these galaxies in redshift space can be reduced and the scope for larger surveys increased in the same way. These approaches would be useful in particular for the forthcoming Fibre Multi Object Spectrograph (FMOS) survey (Dalton et al. 2006) and the planned Wide Field Multi Object Spectrograph (WFMOS) survey.

The outline of this paper is as follows. In Section 2, we present the data used to provide correlations found and examine a low-redshift sample (SDSS) as well as a high-redshift sample (DEEP2). Section 3 presents the methodology used to extract the correlations from the data and describes the found results. We attempt to use LWR and ANNZ for this purpose. In Section 4, we present methodology for the classification of active galactic nuclei (AGN) and star-forming (SF) galaxies. We comment on the prospects that this technique has in speeding up future redshift surveys in Section 5 and conclude in Section 6. We also provide two short appendices describing the mathematics of the LWR and ANNZ methods used.

2 THE DATA

2.1 SDSS

The low-redshift data used in this work are a small sample taken from the SDSS Data Release 2 (Abazajian et al. 2004). Photometry is available for all galaxies in five optical bands ($u, g, r, i$ and $z$). We considered the same flux-limited sample described in Stasińska et al. (2006). This is a random sample of 20000 galaxies with reddening-corrected Petrosian $r$-band magnitudes $r \leq 17.77$, and Petrosian $r$-band half-light surface brightness $\mu_{50} \leq 24.5$ mag arcsec$^{-2}$ (Strauss et al. 2002). As a quality cut, we selected only the objects that show a signal-to-noise ratio greater than 5 in the $g, r$ and $i$ bands. We plot the redshift distribution of this sample in Fig. 1.

The SDSS spectra used here cover the wavelength range 3800–9200 Å, and have a spectral resolution of $\sim 1800$. They were taken with the standard 3 arcsec diameter fibres in the SDSS spectrograph. The spectra are first corrected for Galactic extinction using the maps of Schlegel, Finkbeiner & Davis (1998) and using the extinction law of Cardelli, Clayton & Mathis (1989). They are then brought to the rest frame and resampled from 3400 to 9000 Å in steps of 1 Å with a flux normalization by the median flux in the 4010–4060 Å region. These procedures are necessary to the spectral analysis described in Section 2.3.

2.2 DEEP2

The data sample used at high redshift was the DEEP2 first data release (Davis et al. 2003). The DEEP project is an ongoing project, producing spectroscopy for targets at a redshift range $0.75 < z < 1.5$. The survey uses the DEIMOS spectrograph (Faber et al. 2003) on the Keck II telescope and aims to target around 40 000 galaxies over 3 deg$^2$. The targets are pre-selected from imaging with $B, R$ and $I$ filters taken with the Canada–France–Hawaii Telescope (CFHT) 12 k camera on the CFHT (e.g. Coil et al. 2004). Galaxies are imaged in $B, R$ and $I$ bands and selected to $R_{AB} < 24.1$. Furthermore, a colour cut is applied to pre-select high-redshift objects above a redshift of $z > 0.7$ and only $\simeq 3$ per cent of galaxies above this redshift are rejected by the colour cut. The spectra are taken at moderately high resolution ($R \sim 5000$) and span the range 6300 $\lambda < 9100$ Å. Hence, the doublet [O ii]$\lambda\lambda 3727$ is found from $0.7 < z < 1.4$. This sample contains 4681 objects and its redshift distribution is plotted in Fig. 1.

2.3 Emission-line measurements

In order to measure the emission lines from the SDSS and DEEP2 galaxy spectra, we have used a code to fit them as Gaussian functions, composed of three parameters: width, offset (with respect to the rest-frame central wavelength) and flux. In the case of the SDSS sample, we measured emission lines using the methods...
described in detail by Cid Fernandes et al. (2005) and Mateus et al. (2006). The following emission lines have been measured: \([\text{O} \text{II}], \text{H}\beta, \text{[O} \text{III}], \text{[O} \text{I}], \text{H}z\) and \([\text{N} \text{II}]\). Lines from the same ion are assumed to have the same width and offset, and we consider the following flux ratio constraints: \([\text{O} \text{II}]\lambda 5007/\text{[O} \text{II}]\lambda 4959 = 2.97\) and \([\text{N} \text{II}]\lambda 6584/\text{[N} \text{II}]\lambda 6548 = 3\). To measure the intensities of emission lines, we have to remove the stellar contribution in the continuum at their wavelength range, mainly to account for absorption features in the emission-line regions. This is done by computing for each SDSS galaxy a synthetic stellar spectrum obtained from a linear combination of simple stellar population spectra that fits the observed continuum in the whole spectral range (after the removal of the zones of emission lines and bad pixels). Removing this synthetic spectrum from the observed one leaves us with a pure emission-line residual spectrum from which we can easily measure the emission lines. This method provides a reliable estimate of the stellar absorption in the entire spectrum, including the windows where emission lines are found. This procedure is very important since the regions of some emission lines (mainly the Hz and H\beta Balmer lines) can be contaminated by strong absorption features which may reduce the EW of the lines. For lines which have no absorption, this should make no difference. We also have compared results by integrating over the spectrum and comparing it to the continuum and results do not change much showing that the method is robust.

For the DEEP2 data, only the \([\text{O} \text{II}]\) line provided suitable measurements. Moreover, we have adopted a distinct approach to measure this emission line. Instead of fitting the continuum with a synthetic spectrum, we have performed a polynomial fit of two continuum windows (3653–3713 and 3741–3801 Å) around the line. The emission line was then measured from the continuum-subtracted spectrum through the Gaussian fitting procedure.

### 3 METHODOLOGY: PREDICTING EMISSION FEATURES WITH BROAD-BAND PHOTOMETRY

It is well known that galaxy colours present a good correlation with the emission-line properties of galaxies, such as their EW. We illustrate this in Fig. 2, where the EWs of emission lines measured from the SDSS spectra are plotted as a function of various galaxy colours. Several correlations can be identified between these quantities, as confirmed by the high values of the Spearman rank correlation coefficients obtained for some of them, particularly those involving the Hz and H\beta emission lines. Therefore, it is not only the 4000 Å break colour index that encodes information about the emission features in galaxies (e.g. Mateus et al. 2006), as typical galaxy colours can also be used to infer these properties in a convenient manner, from photometric data only. On this basis, we use here different methods to combine all information available in the galaxy colours to produce an empirical relation between colours and emission strength.

We have analysed the data with two different techniques to reinforce the robustness of the correlations found and to compare the two techniques used. We have used Artificial Neural Networks (Bishop 1995), as implemented in the publicly available ANNz code (Collister & Lahav 2004 and references therein) and a LWR (Atkeson, Moore & Schaal 1997). See the Appendices for the mathematical details and implementation of these methods.

Both of the techniques we present here rely on a training set which is representative of the true population of galaxies in order to retrieve the information on the line features in each galaxy. We have separated the sample into two groups, one group which we have used to train the machine learning algorithms and a second sample which we will name the testing set which is set aside and then used to test the reliability of the methods. We have set aside 1100 galaxies in DEEP and 14 000 galaxies in the SDSS as testing sets.

The rest of the sample was used to train the algorithms, that is 2000 galaxies to train in the DEEP sample and 6000 galaxies to train in the SDSS sample. For both methods used here, we have to subdivide this sample into a training sample and a validation sample. In the case of neural networks, this is done to prevent overfitting. In the case of the LWR method, this is done to obtain the best kernel value \(K\) which is appropriate to the data. Both prescriptions are explained in the Appendix. The architecture of the neural network used is described in the Appendix and is of the type \(N : 2N : 2N : 1\) where \(N\) is the number of magnitudes available which is three in the case of DEEP data and five in the case of SDSS.

With the DEEP sample, the only line which we have produced a fit for was the \([\text{O} \text{II}]\) line. We have performed this analysis only on this line because the DEEP survey relies heavily on the \([\text{O} \text{II}]\) line for redshift estimation, therefore having a high completeness. Other lines do not appear in a large fraction of the galaxies in the entire sample. In the sample from SDSS, we performed this analysis for all measured lines: \([\text{O} \text{II}], \text{H}\beta, [\text{O} \text{III}], [\text{O} \text{I}], \text{H}z\) and \([\text{N} \text{II}].\)

In Fig. 3 we plot our attempts to recover the EW of the \([\text{O} \text{II}]\) line in the DEEP survey using broad-band photometry only. We have plotted a scatter plot of the real EW versus the predicted EW for the testing set given the training set. As we can see, the logarithm of the EW is predicted reasonably well and show that this weak correlation can be inferred very well with the techniques we have proposed here. We can predict the log of the EW with an error of 0.27 without any spectroscopic information on the redshift of each galaxy.

We present in Fig. 4 the results we find by fitting different line widths to SDSS galaxies. Again, in these plots we have used only galaxies in the testing set which have not been used in the training process. We find that some line EWs are relatively well predicted by galaxy magnitudes and colours. This is the case for the hydrogen lines where a reasonably strong correlation was found. We have found, however, that the collisional lines observed are harder to predict. For instance, the \([\text{O} \text{I}]\) line EW had virtually no correlation with any combination of the colours and magnitudes found.

We compare results for both of the techniques using the same testing sets in Fig. 4. We plot them side by side so that we can see the relative comparison between the neural network non-linear fit to the data and the locally linear regression used. We have inspected the data and found that there was no significant difference between both methods. Both predictions have yielded results with a very similar scatter given the data volume available. We conclude that for fitting purposes neural networks and a LWR perform well and on an equal footing.

We have also attempted to simply classify objects into line-emitting and non-line-emitting. We have made this separation for each galaxy upon the simple criterion: whether an emission line is detected above a certain value. This separation has been done for each line emission analysed. For instance, if a certain galaxy has an \([\text{O} \text{I}]\) line but does not possess an Hz line, it is considered as an emitting object for the purpose of the analysis on the \([\text{O} \text{I}]\) line but...
Figure 2. EWs of the emission lines [O II], Hβ, [O III], Hα and [N II] measured from the SDSS spectra as a function of the colours (u-r), (g-r), (u-i), and (g-i). The top-right numbers are the total number of galaxies in each panel and the Spearman rank correlation coefficient. The grey-scale level represents the number of galaxies in each pixel and darker pixels having more galaxies.
Spectral features from broad-band photometry

Figure 3. The scatter plot for the 1100 galaxies that were used as a testing set for the prediction of the EW of the [O II] line in the DEEP sample. The artificial neural network method used in this plot. We can see a correlation in the log of the predicted EW and the actual measurement. The scatter in the log is of 0.27. Similar results with virtually the same scatter are found with the LWR technique.

it is considered as a non-emitting object for the purposes of the Hα line analysis.

We have then used the training set in the following way. We have assigned the value of 1 to emitting objects and the value of 0 to ‘non-emitting’ objects, defined by the line being detected in our code above a signal-to-noise ratio of 3. A neural network was then trained in this specific training set. When a neural network is trained by binary values between 0 and 1, it returns a probability (e.g. Lahav et al. 1996 and references therein) of that given object belonging to the class described by the binary number. So, in our case the value returned by the neural network will be the probability of that object having the emission feature tested for. We have also attempted to perform the same analysis with the LWR technique but found that despite the LWR technique producing predictive results which are comparable and as good as the neural network fit, it was not able to perform this classification task nearly as well as the neural networks. This was to be expected as the LWR method is not directly applicable to classification tasks. This is because the linear fit results are not representative of a probability range in the same way the neural networks are.

We plot these results in Fig. 5. We can see that the results are varied and depend on the emission line being assessed. For instance, there is a good prospect of training a neural network to distinguish between hydrogen emitters and [N II] emitters at low redshift, but on the other hand the [O I] line exhibits a very poor correlation between colours and magnitudes and its EW. The other oxygen lines exhibit some correlation that are picked up by our methods.

In order to have an idea of what contamination would be introduced but performing a selection based on the predicted EW of a galaxy with our method, we have plotted histograms for the entire sample and then chosen the centre of the distribution of the EWs in log space. We have then plotted the histograms for the samples with high- and low-predicted EW in Fig. 6.

While predicting the EWs for these emission lines we have found the already known result of galaxy bi-modality. We know that galaxies are bimodal and that late-type galaxies have higher star formation.

Figure 4. The scatter plot of the log of the real EW versus the predicted EW of galaxies for different lines. For all the plots, we have used only galaxies in the testing set which have not made part on the training of the machine learning algorithm used. We present here results using both methods described in this paper. As we can see, the results found with both results are comparable and we found no evidence that one method performed better than the other given the data volume available. We can also see that some lines present a better correlation than others and that generally recombination lines are better predicted than collisional lines.

rates and therefore stronger emission features whereas the opposite is true for late-type galaxies. We can see this bi-modality in the histogram of the EWs for the different lines in Fig. 6. This is not found on every set of lines because there are selection effects (fainter lines are simply not seen in some galaxies). For instance, the ratio of the Hα and Hβ lines for a single galaxy should be roughly constant just below 3. This is because both lines are recombination lines and the temperature dependence of the emissivity is roughly the same for both lines. However, we do not find this bimodality for Hβ but we find it for Hα. This is simply because the Hβ flux is smaller and this makes so that a lot of the fainter lines remain undetected. Furthermore, we argue that the correlations found in Hβ are strong which suggests that the method we are using is going beyond simply separating the bi-modality and is classifying the line widths according to their size.

4 METHODOLOGY: AGN/STAR FORMATION CLASSIFICATION

4.1 Traditional spectral galaxy classification

We adopted a traditional procedure to classify galaxies according to their emission-line properties. By examining diagnostic diagrams formed by line ratios of optical emission lines, such as the [O III]/Hβ versus [N II]/Hα diagram proposed by Baldwin, Phillips & Terlevich (1981), we can distinguish emission-line galaxies according to the mechanism responsible for producing the lines. In such diagrams, hosts of AGN and SF galaxies form two very distinct branches, or wings, making the task to separate them easier. We use this ‘convenient’ diagram to classify galaxies following the classification scheme discussed in Stasińska et al. (2006) based on a theoretical curve used to distinguish galaxies with pure star formation from those objects with some contribution from nuclear activity to the line intensities. Galaxies below such a curve are classified as normal SF galaxies. In addition, we have also used the empirical curve proposed by Kauffmann et al. (2003) to identify the AGN hosts. Galaxies between the two curves are hybrid objects (the contribution of the AGN to the Hβ emission of galaxies below the Kauffmann et al. line is at the most 3 per cent) and therefore will not be included in our analysis. The AGN hosts are then selected as those objects located above the Kauffmann et al. empirical curve.

4.2 Automated spectral galaxy classification

In this section, we attempt to classify the spectral features instead of simply predicting the EW of lines. We have described in Section 4.1 how we separate galaxies with emission features into AGN or SF galaxies.

We have separated the 20,000 SDSS galaxies we used in this paper into the following classes: passive galaxies, AGN, SF galaxies and hybrid objects. Some objects were not classified by the code analysing the spectroscopic data and we have removed those
Spectral features from broad-band photometry

We have shown that it is possible to predict emission features using broad-band photometry. We find that the correlations vary in strength from line to line. Particularly, recombination lines are more predictable than collisional lines. We have not included any information on the shape or size of the galaxies. We expect that the results found here would be stronger if information such as shapelet coefficients for galaxies is included into the analysis as we know that galaxies with different star formation rates and therefore different line strengths have different morphologies.

We argue that this technique can be used in future spectroscopic redshift surveys to speed them up. Future facilities, such as FMOS or WFMOS, will produce spectroscopic surveys in the optical and the infrared part of the spectrum targeting mainly the [O\textsc{ii}] and the H\textbeta lines. We have shown here that there is a reasonable predictability of the [O\textsc{ii}] line at high redshift and that the H\textalpha and H\beta lines are well predicted in the low-redshift Universe. A high-redshift complete sample of around 10,000 objects could be used to train a network to then select the preferred targets.

If one chooses to predict the line strengths, one alternative would be to fit stellar population synthesis models to the observed spectrum and from that infer a star formation rate and therefore a line flux. We argue that the technique we have proposed here is complementary and one would be able to encode information on the morphology of the galaxy easily whereas if one fits stellar population models to the training data it would be hard to include special information in the analysis.

6 CONCLUSION

We have looked, in this paper, for empirical relations between the EW for several different lines and the broad-band colours for the same objects. We used an automated way to explore the correlations found in the data via the use of LWR and ANN. The two methods give very similar results.

We have performed the analysis in two samples, a low-redshift sample obtained from the SDSS covering the nearby Universe and a high-redshift sample from the first data release of the DEEP survey covering a deeper sample off the Universe. We found that in the DEEP data one could predict the EW of the [O\textsc{ii}] line relatively well by looking at broad-band colours only. With the SDSS data, six lines were predicted from a training set sample. In general, collisional lines presented little correlation although some of them were reasonably predicted by the colours. There was a stronger correlation found in recombination lines.

We have compared the power of prediction of both methods used. We have concluded that both Artificial Neural Networks and the LWR methods are capable of recovering most of the information encoded in the training set with little statistical difference between both methods. We have used both the methods for classification purposes and found that the Neural Networks are capable of classifying the objects into line-emitting and non-line-emitting but the LWR method was unable to do so. We have shown that it is possible to classify galaxies into AGN, passive galaxies and line-emitting galaxies well without galaxy spectra, using solely colours and a training set of the order of 15,000 galaxies.

We have discussed the prospects of speeding up redshift surveys with this and we conclude that with a reasonable training set of the order of 10,000 galaxies one would be able to considerably speed up future surveys done with instruments such as FMOS and WFMOS. Furthermore, in this paper we have only taken advantage of the colour data; however the methods described can very easily accommodate other information such as morphology or size.

ACKNOWLEDGMENTS

We acknowledge Manda Banerji and Eduardo Cypriano for useful comments as well as Richard Ellis, Peder Norberg, John Peacock and other members of one of the WFMOS design study team. FBA acknowledges the support of the Leverhulme Trust via an Early
LWR is a machine learning method capable of mapping complex, non-linear relations between variables (Atkeson et al. 1997).

Like any other global fits, LWR minimizes a quantity related to the chi-square. The difference is that each data point of the training set is given a corresponding weight, which depends on a given query point. Thus, the fitting parameters are valid only locally for that particular query point. In this method, any data available for the machine learning would be separated into a training set and a validation set.

We use a particular species of ANN known formally as a multi-layer perceptron (MLP). A MLP consists of a number of layers of notation we arrange these in a vector $\beta$:

$$\beta = (X^T X)^{-1} X^T y$$

where $x_i$, $y_i$ are the magnitudes, background (i.e. $t_1 = 1, t_2 = x_1, t_3 = x_1^2$ and so forth). This equation can be written as

$$f(x) = \beta^T t(x)$$

where $t(x)$ is the vector of polynomial terms. Here, the weights are recomputed according the distance between points $x_i$ and the query point and the matrix $\beta$ is computed according to

$$\beta = (X^T X)^{-1} X^T y$$

and

$$f(x) = \beta^T t(x)$$

hence we can solve for the best solution by using a Cholesky decomposition (Press et al. 1992). Now that the $\beta$ vector and the kernel are chosen, we simply apply the relation using the testing set as the query points.

APPENDIX B: ARTIFICIAL NEURAL NETWORKS

We use a particular species of ANN known formally as a multi-layer perceptron (MLP). A MLP consists of a number of layers of nodes (Fig. B1; see e.g. Bishop 1995 and references therein for background). The first layer contains the inputs, which in this paper are the magnitudes, $m_i$, of a galaxy in a number of filters (for ease of notation we arrange these in a vector $m = (m_1, m_2, \ldots, m_n)$). The final layer contains the outputs; here the EW or emission probability. Intervening layers are described as hidden and there is complete freedom over the number and size of hidden layers used. The nodes in a given layer are connected to all the nodes in adjacent layers. A particular network architecture may be denoted by $N_{in}; N_1; N_2; \ldots; N_{out}$, where $N_{in}$ is the number of input nodes, $N_1$ is the number of nodes in the first hidden layer and so on. For example, 9:6:1 takes nine inputs, has six nodes in a single hidden layer and gives a single output.
Spectral features from broad-band photometry

Figure B1. A schematic diagram of a MLP, as implemented by ANNz, with input nodes taking, for example, magnitudes $m_i = -2.5 \log_{10} f_i$ in various filters, a single hidden layer and a single output node. The architecture is $n:p:1$ in the notation used. Each connecting line carries a weight $w_{ij}$. The bias node allows for an additive constant in the network function defined at each node. More complex networks can have additional hidden layers and/or outputs. Here, the EW takes the same role as the redshift.

Each connection carries a weight, $w_{ij}$; these comprise the vector of coefficients, $w$, which are to be optimised. An activation function, $g_j(u_j)$, is defined at each node, taking as its argument

$$u_j = \sum_i w_{ij} g_i(u_i),$$

where the sum is over all nodes $i$ sending connections to node $j$. The activation functions are typically taken (in analogy to biological neurons) to be sigmoid functions such as $g_j(u_j) = 1/[1 + \exp(-u_j)]$, and we follow this approach here. An extra input node – the bias node – is automatically included to allow for additive constants in these functions.

For a particular input vector, the output vector of the network is determined by progressing sequentially through the network layers, from inputs to outputs, calculating the activation of each node (hence this type of neural network is often referred to as a feed-forward network).

Given a suitable training set of galaxies for which we have both photometry, $m$, and here the EW, $EW_{\text{train}}$, the ANN is trained by minimizing the cost function

$$E = \sum_i (EW_{\text{out}}(w, m_i) - EW_{\text{train},i})^2,$$  \hspace{1cm} (B2)

with respect to the weights, $w$, where $EW_{\text{out}}(w, m_i)$ is the network output for the given input and weight vectors, and the sum is over the galaxies in the training set. To ensure that the weights are regularized (i.e. they do not become too large), an extra quadratic cost term

$$E_w = \beta \sum_{i,j} w_{ij}^2$$  \hspace{1cm} (B3)

is added to equation (B2).

We use an iterative quasi-Newton method to perform this minimization. Details of the minimization algorithm and regularization may be found in Bishop (1995) and Lahav et al. (1996, appendices).

After each training iteration, the cost function is also evaluated on a separate validation set. After a chosen number of training iterations, training terminates and the final weights chosen for the ANN are those from the iteration at which the cost function is minimal on the validation set. This is useful to avoid overfitting to the training set if the training set is small. The trained network may then be presented with previously unseen input vectors, and the outputs computed.

To implement the ANNs code to our problem all that is needed is to regard the output node as the EW instead of as the redshift. In this work, we found it optimal to work with the log(EW).

2 http://zuserver2.star.ucl.ac.uk/~lahav/annz.html

This paper has been typeset from a TeX/EBX file prepared by the author.