Estimating photometric redshifts with artificial neural networks

Andrew E. Firth,[1] Ofer Lahav[1]★ and Rachel S. Somerville[1,2]

[1]*Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA*
[2]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109-1090, USA*

ABSTRACT

A new approach to estimating photometric redshifts – using artificial neural networks (ANNs) – is investigated. Unlike the standard template-fitting photometric redshift technique, a large spectroscopically-identified training set is required but, where one is available, ANNs produce photometric redshift accuracies at least as good as and often better than the template-fitting method. The Bayesian priors on the underlying redshift distribution are automatically taken into account. Furthermore, inputs other than galaxy colours – such as morphology, angular size and surface brightness – may be easily incorporated, and their utility assessed.

Different ANN architectures are tested on a semi-analytic model galaxy catalogue and the results are compared with the template-fitting method. Finally, the method is tested on a sample of $\sim$20 000 galaxies from the Sloan Digital Sky Survey. The rms redshift error in the range $z \lesssim 0.35$ is $\sigma_z \sim 0.021$.

Key words: methods: data analysis – surveys – galaxies: distances and redshifts.

1 INTRODUCTION

The basic photometric redshift technique is to use the colours of a galaxy in a selection of medium- or broad-band filters as a crude approximation of the spectral energy distribution (SED) of the galaxy, in order to find its redshift and spectral type. The technique is very efficient compared with spectroscopic redshifts since the signal-to-noise in broad-band filters is much greater than the signal-to-noise in a dispersed spectrum and, furthermore, a whole field of galaxies may be imaged at once while spectroscopy is limited to individual galaxies or those that can be positioned on slits or fibres. However, photometric redshifts are only approximate at best and are sometimes subject to complete misidentifications. For many applications, though, large sample sizes are more important than precise redshifts and photometric redshifts may be used to good effect.

Photometric redshifts date back to Baum (1962, see also Hogg et al. 1998; Weymann et al. 1999). They have been used extensively in recent years on the ultra-deep and well-calibrated Hubble Deep Field observations (e.g. Gwyn & Hartwick 1996; Connolly, Szalay & Brunner 1998; Fernández-Soto, Lanzetta & Yahil 1999; Fontana et al. 2000; Fernández-Soto et al. 2001; Massarotti, Iovino & Buzzoni 2001a; Massarotti et al. 2001b). The most commonly used approach is the template-fitting technique. This involves compiling a library of template spectra – either theoretical SEDs from population synthesis models (e.g. GISSEL – Bruzual & Charlot 1993) or empirical SEDs (e.g. Coleman, Wu & Weedman 1980, hereafter CWW). Then the expected flux through each survey filter is calculated for each template SED on a grid of redshifts, with corrections

for ISM, IGM and Galactic extinction where necessary. A redshift and spectral type are estimated for each observed galaxy by minimizing χ^2 with respect to redshift, z, and spectral type, SED, where

$$\chi^2(z, \text{SED}) = \sum_i \left[\frac{f_i - \alpha(z, \text{SED}) t_i(z, \text{SED})}{\sigma_i} \right]^2, \quad (1)$$

f_i is the observed flux in filter i, σ_i is the error in f_i, $t_i(z, \text{SED})$ is the flux in filter i for the template SED at redshift z, and $\alpha(z, \text{SED})$ (the scaling factor normalizing the template to the observed flux) is determined by minimizing equation (1) with respect to α, giving

$$\alpha(z, \text{SED}) = \left[\sum_i \frac{f_i t_i(z, \text{SED})}{\sigma_i^2} \right] \Big/ \left[\sum_i \frac{t_i(z, \text{SED})^2}{\sigma_i^2} \right]. \quad (2)$$

The template-fitting photometric redshift technique makes use of the available and reasonably detailed knowledge of galaxy SEDs and in principle it may be used reliably even for populations of galaxies for which there are few or no spectroscopically confirmed redshifts. However, crucial to its success is the compilation of a library of accurate and representative template SEDs (see e.g. Hogg et al. 1998; Firth 2002b). Empirical templates are typically derived from nearby bright galaxies, which may not be truly representative of high redshift galaxies. Conversely, while theoretical SEDs can cover a large range of star formation histories, metallicities, dust extinction models, etc., not all combinations of these parameters (at any particular redshift) are realistic, and the ad hoc inclusion of superfluous templates increases the potential for misidentifications when using observations with noisy photometry.

An alternative approach can be used when one has a sufficiently large (e.g. $\sim$100–1000, depending on the redshift range) and

representative subsample with spectroscopic redshifts. Then one can fit a polynomial, or other function, mapping the photometric data to the known redshifts and use this to estimate redshifts for the remainder of the sample with unknown redshifts (e.g. Connolly et al. 1995b; Brunner, Szalay & Connolly 2000; Sowards-Emmerd et al. 2000). With this approach, errors in the estimated redshifts may also be estimated analytically or via Monte Carlo simulations.

An extension of the latter approach is to use artificial neural networks (ANNs), which have been used before in astronomy for, amongst other things, galaxy morphological classification (e.g. Storrie-Lombardi et al. 1992; Naim et al. 1995; Lahav et al. 1996), morphological star/galaxy separation (e.g. Bertin & Arnouts 1996; Andreon et al. 2000) and stellar spectral classification (e.g. Bailer-Jones, Irwin & von Hippel 1998; Allende Prieto et al. 2000; Weaver 2000). Essentially, an ANN takes a set of inputs (e.g. logarithms of fluxes – i.e. magnitudes – in different filters) for each object, applies some non-linear function, and outputs a value (e.g. the estimated redshift). The ANN is first trained – i.e. the coefficients (weights) of the function are optimized – by using a training set where the desired output is known. The ANN may then be used on any number of other objects with similar inputs (i.e. magnitudes in the same filter set) but unknown outputs (i.e. redshifts).

As well as using all of the information contained in the magnitudes and colours, provided the training set is a representative subsample of the data, the ANN will also take into account the Bayesian priors on the galaxy redshift distribution (cf. Benitez 1998; Teplitz et al. 2001). While choosing a template library that is both sufficient and non-superfluous is a source of concern for the template-fitting method, ANNs automatically fit the true range of galaxy SEDs. Another potential advantage of ANNs relative to the template-fitting method is that the weights applied to each filter may be more optimal than simple χ^2-weighting. In addition, one can also feed in other observational input such as image size or surface brightness, morphology and concentration parameters where such data are available. It is interesting then to see how the two methods compare.

This paper explores the use of ANNs as a potential tool for photometric redshift determination. The layout of this paper is as follows. In Section 2 the ANNs are described and in Section 3 a semi-analytic model (used to provide a simulated galaxy catalogue) is introduced. In Section 4 the ANN parameters (architecture and training set size) are investigated using the simulated galaxy catalogue and in Section 5 the performance of ANNs are compared with the performance of the traditional template-fitting method. Section 6 looks at the effect of photometric noise and in Section 7 ANNs are investigated as a method for also determining spectral type from redshifted data. In Section 8, ANNs are tested on Sloan Digital Sky Survey observational data. The science prospects are briefly discussed in Section 9.

2 ARTIFICIAL NEURAL NETWORKS

An ANN comprises a set of input nodes, one or more output nodes, and one or more hidden layers each containing a number of nodes (Fig. 1; see e.g. Bishop 1995, and references therein, for background). 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 9 inputs, has 6 nodes in a single hidden layer and gives a single output. The nodes are connected and each connection carries a weight comprising the vector of coefficients w that are to be optimized. Unless otherwise stated, every node is here assumed to be connected to every node in the previous layer and to every node
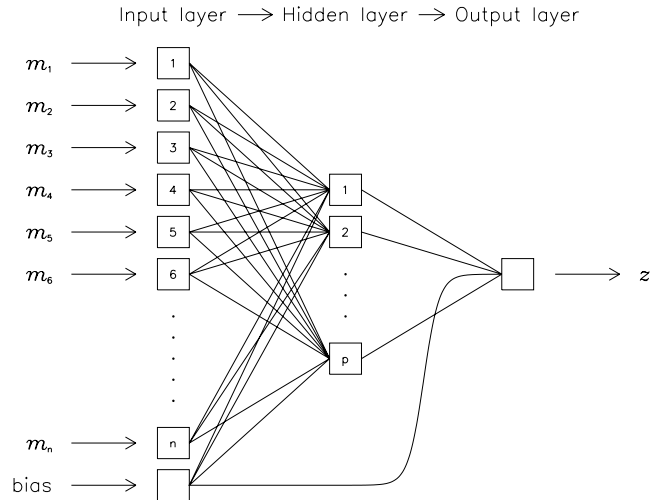

Figure 1. A schematic diagram of an ANN with input nodes taking, for example, magnitudes $m_i = -2.5 \log_{10} f_i$ in various filters, a single hidden layer, and a single output node giving, for example, redshift z. The architecture is $n : p : 1$ in the notation used in this paper. Each connecting line carries a weight w_j. The bias node allows for an additive constant in the network function defined at each node. More complex nets can have additional hidden layers.

in the next layer only, but it is certainly possible to have more or less interconnected nets. The input parameters for each object are represented by the vector x (e.g. the magnitudes in a set of filters). Given a training set of inputs x_k and desired outputs z_k (e.g. the redshifts), the ANN is optimized by minimizing the cost function

$$E = \frac{1}{2} \sum_k [z_k - F(w, x_k)]^2. \tag{3}$$

The function $F(w, x_k)$ is given by the network. A function g_p is defined at each node p, taking as its argument

$$u_p = \sum_j w_j x_j, \tag{4}$$

where the sum is over the input nodes to p. These functions are typically taken (by analogy with biological neurons) to be sigmoid functions such as $g_p(u_p) = 1/[1 + \exp(-u_p)]$ (used here). An extra input node – the bias node – is automatically included to allow for additive constants in these functions. The combination of these functions over all the network nodes makes up the function F. A programme kindly provided by B. D. Ripley was used to train the networks. The programme takes as its input a network architecture, a training set and a random seed to initiate the weight vector, w,[1] and uses an iterative quasi-Newton method (see e.g. Bishop 1995) to minimize the cost function. To ensure that the weights are regularized (i.e. that they do not become too large), an extra quadratic cost term

$$E_w = \beta \frac{1}{2} \sum_j w_j^2, \tag{5}$$

was added to equation (3). A value of $\beta = 0.0001$ was chosen empirically to optimize the ANN performance. After each training iteration, the cost function is evaluated on a separate validation set.

---

[1] The initial weights were randomly chosen from a uniform distribution with range $[-0.7, 0.7]$.

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 over-fitting to the training set if the training set is small.

3 MODEL GALAXY CATALOGUES

3.1 Semi-analytic models

To provide a galaxy catalogue on which to train and test ANNs, a semi-analytic model was used. Semi-analytic models are an attempt to use simple recipes to parametrize the main physical processes of galaxy formation within the hierarchical paradigm of galaxy formation (see, for example, Kauffmann, White & Guiderdoni 1933; Cole et al. 1994). In these models, Monte Carlo techniques may be used efficiently to generate large mock galaxy catalogues with a (broadly-speaking) realistic distribution of galaxy types, luminosities, colours and redshifts. Here, the current version of the code developed by Somerville (1997) is used. This has been shown to produce good agreement with many properties of local and high-redshift galaxies (Somerville & Primack 1999; Somerville, Primack & Faber 2001; Firth et al. 2002a).

In this model (see Somerville & Primack 1999, for details), the number density of haloes of various masses at a given redshift is determined by an improved version of the Press–Schechter model (Sheth & Tormen 1999) and the formation and merging of dark matter haloes as a function of time is represented by a 'merger tree'. The cooling of gas, formation of stars, and reheating and ejection of gas by supernovae within these haloes are modelled by simple recipes. Cold gas is assumed to cool initially into, and form stars within, a rotationally supported disc. Major mergers between galaxies destroy the discs and create spheroids. Galaxy mergers also produce bursts of star formation. The chemical evolution and star formation history of each galaxy is traced and convolved with multi-metallicity stellar population synthesis models (Devriendt, Guiderdoni & Sadat 1999), and a dust extinction law, in order to calculate the galaxy's SED.

There are several advantages in using a semi-analytic model here. First, an arbitrarily large number of galaxies may be generated over any desired redshift range, and with any magnitude limit. The 'true' redshift and magnitudes (prior to the addition of photometric noise) are known precisely. At present there is no large observed spectroscopic sample at high redshift, and those spectroscopic samples that do exist tend to be biased towards luminous galaxies with prominent emission lines at optical wavelengths. On the other hand, simpler model galaxy catalogues (e.g. PLE models) are less likely to produce realistic distributions of galaxy SEDs in terms of composite stellar populations, ages, metallicities and the effects of dust (all as a function of redshift). Where (when) suitable spectroscopic samples exist, the model catalogues could be replaced by observed photometric and spectroscopic samples and the results would be expected to be comparable (see, for example, Section 8).

3.2 Preparing the input catalogue

An $H < 22$ catalogue in $UBVRIH$ was generated using the semi-analytic model. To simulate a real galaxy survey, photometric noise was added to this catalogue, simulating 5-σ magnitude limits of $U = 25.1$, $B = 26.6$, $V = 26.1$, $R = 25.6$, $I = 24.7$ and $H = 20.5$ (typical of current and future large surveys aimed at studying large-scale structure at high redshifts – e.g. the LCIR Survey, Firth et al. 2002a).

An extra 0.05 mag rms error term was included to simulate seeing variations, zero-point inaccuracies and other sources of photometric errors. Finally, an $H < 20.5$ 'noisy' sample was drawn from this catalogue.

So that all weights are treated fairly equally in equation (5), it is useful to normalize the magnitudes in each filter and the model redshifts to the range [0, 1]. The exact form of the normalization is unimportant provided the same normalization is used for both training and using the ANN. For definiteness, in each filter the mean magnitude (derived from an $H < 20.5$ 'noiseless' sample) was subtracted, and the range [−5, 5] was mapped linearly to [0, 1]. Furthermore, the redshift range [0, 3.5] was mapped linearly to [0, 1].

4 SELECTING NETWORK PARAMETERS

4.1 Number of training iterations

First of all, the required number of training iterations, N_{iter}, was investigated. Clearly this will depend on the characteristics of the data set and the complexity of the network architecture. There is also an element of chance due to the randomized initial weights. After an ANN has been trained, its performance is assessed by running a *testing* set through it (distinct from the *training* and *validation* sets) and calculating the rms of $\Delta_z = z_{model} - z_{phot}$, where z_{model} are the model redshifts and z_{phot} are the corresponding ANN photometric redshifts. Fig. 2 plots the change in rms as N_{iter} is increased. Three architectures, covering a range in complexity, are compared and, for each architecture, five ANNs were produced starting with different random seeds. A training set of size 10 000 was used (cf. Section 4.3) and the ANNs were tested on a separate testing set, also of size
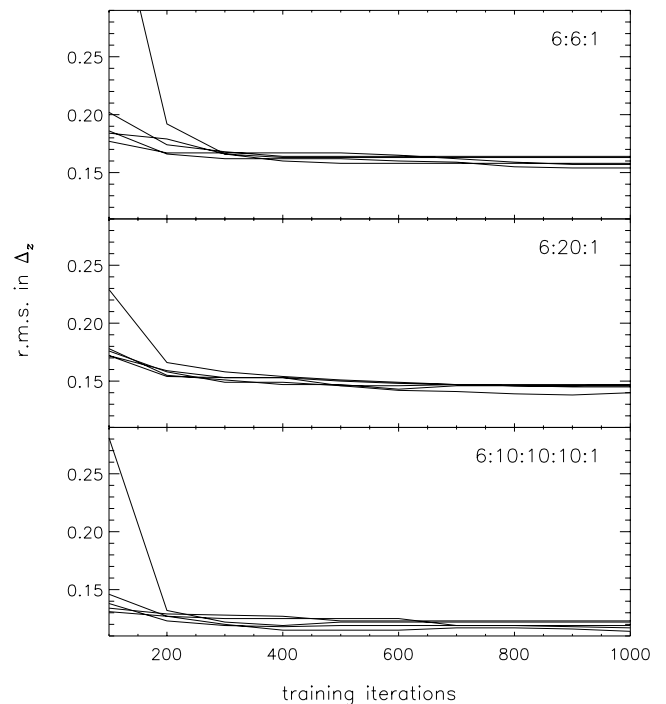

Figure 2. The rms of $\Delta_z = z_{model} - z_{phot}$ (measured on 10 000 test galaxies) as a function of the number of training iterations for three ANN architectures (indicated at the upper right of each panel). For each architecture, five ANNs were trained, each initialized with a different random seed. A training set of size 10 000 was used.
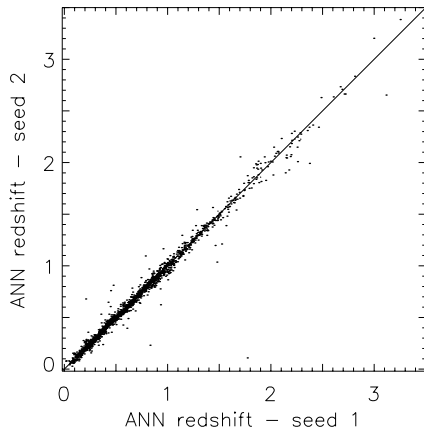
Figure 3. Comparison of photometric redshifts (using 2000 test SAM galaxies) for two ANNs initialized with different random seeds. A $6:10:10:10:1$ network architecture, 10 000 training galaxies and 1000 training iterations were used. The rms of $\Delta_z = z_1 - z_2$ is 0.065.

10 000. Note that the testing set could be any size. A sample of size 10 000 was chosen simply to provide good statistics.

The rms decreases as N_{iter} is increased but levels off for large N_{iter}. For most random seeds, improvement beyond $N_{iter} \sim 500$ is slow. For the remainder of this paper, in which a similar range of architectures, and the same – or simpler – data sets are considered, N_{iter} will be restricted to 1000.

Fig. 3 compares the estimated photometric redshifts on the testing set, using the $6:10:10:10:1$ architecture, for two different random seeds after 1000 training iterations. The two ANNs closely agree. Generally, different initial random seeds lead to ANNs with similar rms accuracy, though there are still differences at the $\lesssim 10$ per cent level. It is useful, therefore, to generate several ANNs using different random seeds. One may then use the validation set to choose the 'best' ANN. A better approach (see e.g. Bishop 1995 for details) is to combine a set of ANNs generated using different random seeds. This is called a 'committee of ANNs'. In this paper, the estimated redshift for each galaxy presented to the committee is taken to be the median of the estimates provided by the individual ANNs. Typically, the committee gives more accurate redshift estimates than any of its
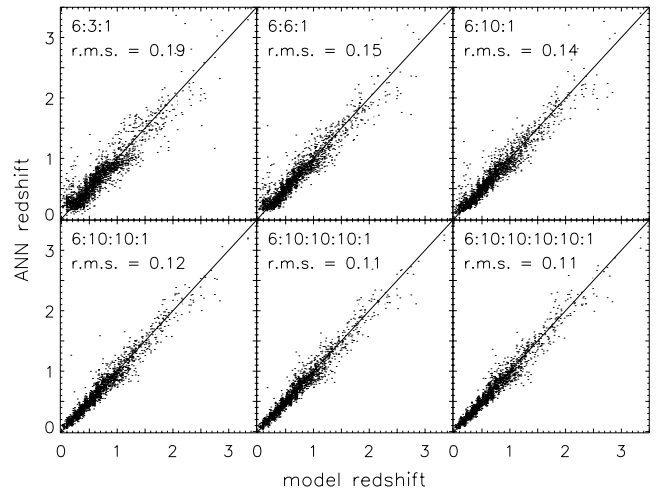

Figure 4. Photometric redshift versus model redshift comparisons (using 2000 test galaxies) for several network architectures (indicated at the upper left of each panel). The simulated SAM catalogue is $H < 20.5$ with 5-σ limits $U = 25.1$, $B = 26.6$, $V = 26.1$, $R = 25.6$, $I = 24.7$ and $H = 20.5$. A committee of 5 ANNs (for each architecture) and a training set of 10 000 galaxies were used. Note the increased scatter at high-redshift where there are fewer training galaxies. The quantization at 0.1 intervals in z is due to poor redshift-space resolution in the colour grid used to determine galaxy colours in the semi-analytic model.

component ANNs taken individually (see Section 4.2, Table 1 for examples).

4.2 Network architecture

More complex network architectures have more free parameters (weights) and therefore allow a closer fit to the data. In any real data set there will be a fundamental limit to the rms fit (due to random noise in the input measurements), and further increases in architecture complexity will provide no significant improvement. In addition, architectures with more weights take longer to train. In any given situation one would like to use the simplest network possible while still obtaining optimal results.

Table 1 compares the rms of $\Delta_z = z_{model} - z_{phot}$, evaluated on a testing set of size 10 000, for several architectures (see also Fig. 4).

Table 1. The mean and rms of $\Delta_z = z_{model} - z_{phot}$ for various network architectures. The simulated SAM catalogue is $H < 20.5$ with 5-σ limits $U = 25.1$, $B = 26.6$, $V = 26.1$, $R = 25.6$, $I = 24.7$ and $H = 20.5$. A training set of 10 000 galaxies was used. Columns 3 and 4 display mean values for five ANNs initialized with different random seeds. Columns 5 and 6 display values for a single committee comprising five ANNs (initialized with different random seeds). Note that a committee produces significantly better results than its component ANNs.

Network architecture	$N_{weights}$	Individual ANN		Committee	
		mean Δ_z	rms of Δ_z	mean Δ_z	rms of Δ_z
6:3:1	25	−0.005	0.192	−0.004	0.187
6:6:1	49	−0.005	0.159	−0.004	0.151
6:10:1	81	−0.004	0.148	−0.005	0.142
6:15:1	121	−0.003	0.146	−0.004	0.140
6:20:1	161	−0.004	0.145	−0.004	0.140
6:30:1	241	−0.003	0.140	−0.003	0.136
6:40:1	321	−0.003	0.143	−0.003	0.139
6:6:6:1	91	−0.002	0.133	−0.002	0.124
6:6:6:6:1	133	−0.001	0.123	−0.000	0.116
6:10:10:1	191	−0.001	0.124	−0.000	0.116
6:10:10:10:1	301	−0.000	0.119	0.000	0.113
6:10:10:10:10:1	411	0.000	0.119	0.000	0.113

For committees of five ANNs, the networks with a single hidden layer (i.e. $6:m:1$) reach a limiting rms of ~ 0.14 (e.g. for $m \sim 15$). There is no significant improvement for larger m. However, adding extra hidden layers offers some improvement, even when the total number of weights is not increased. The committees of five $6:6:6:6:1$, $6:10:10:1$ and $6:10:10:10:1$ ANNs in Table 1 all produce rms values of ~ 0.115. The scatter in the rms between different committees is of order ± 0.002 for these architectures, and increasing the number of ANNs in the committee leads only to a small improvement in the rms (e.g. ~ 0.001 decrease in the rms for a committee of 25 ANNs). Adding further hidden layers (e.g. $6:10:10:10:10:1$) leads to little or no further improvement.

For the remainder of this paper, a $6:6:6:6:1$ architecture will be used as the fiducial ANN, since this architecture gives results comparable with the best results in Table 1 but has relatively few weights. Unless stated otherwise, redshifts will be estimated using a committee of five such ANNs – taking the median redshift estimate of five for each galaxy.

Clearly, the required network complexity depends on the data set. For data with a higher signal-to-noise ratio, the fundamental rms limit on Δ_z will be lower and a more complex architecture may be necessary to reach this limit. Conversely, for data covering a smaller redshift range, fewer free parameters will be necessary to model the mapping from colours to redshift, and a simpler architecture may suffice.

4.3 Size of training set

Often one will have no choice concerning the size of the training set, N_{train}. However, when designing surveys, it is useful to assess what size of training set is necessary to provide a given redshift accuracy. Fig. 5 plots the rms of $\Delta_z = z_{\text{model}} - z_{\text{phot}}$, evaluated on a testing set of size 10 000, as a function of N_{train}, for the $6:6:6:6:1$ architecture (using a single random seed). The rms decreases as N_{train} increases but begins to level off for large N_{train} values. Clearly there is a trade-off – one can obtain fairly good results for quite small training sets (e.g. rms ~ 0.15 for $N_{\text{train}} = 1000$) but larger training sets can give significant improvements (e.g. rms ~ 0.12 for $N_{\text{train}} = 10\,000$).
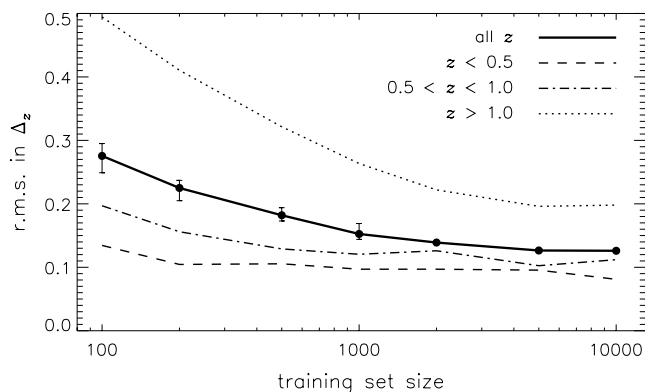

Figure 5. Photometric redshift accuracy as a function of training set size N_{train}. The plot shows the rms of $\Delta_z = z_{\text{model}} - z_{\text{phot}}$, evaluated on 10 000 test galaxies. The simulated SAM catalogue is $H < 20.5$ with 5-σ limits $U = 25.1$, $B = 26.6$, $V = 26.1$, $R = 25.6$, $I = 24.7$ and $H = 20.5$. A $6:6:6:6:1$ network architecture (with a single random seed) was used. For $N_{\text{train}} < 10\,000$, up to 10 ANNs were generated using separate training sets; points and error bars show, respectively, medians and interquartile ranges for the different ANNs.
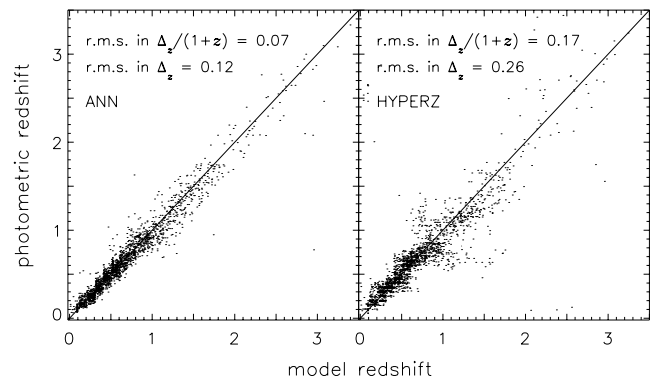

Figure 6. A comparison of ANN and HYPERZ photometric redshifts with model redshifts on 2000 test galaxies. The simulated SAM catalogue is $H < 20.5$ with 5-σ limits $U = 25.1$, $B = 26.6$, $V = 26.1$, $R = 25.6$, $I = 24.7$ and $H = 20.5$. A committee of five $6:6:6:6:1$ ANNs and a training set of 10 000 galaxies were used. Eight Bruzual & Charlot GISSEL'98 evolving SEDs were used as templates in HYPERZ with A_V in the range 0.0–1.2.

Also plotted in Fig. 5 are the rms values for different redshift bins: $z < 0.5$, $0.5 < z < 1.0$ and $z > 1.0$. In the testing set, galaxies are distributed between these bins in the approximate ratio $2:2:1$. Since there are fewer training galaxies in the high-redshift bin, the network weights are less constrained and the rms values are larger. From $N_{\text{train}} = 200$ to $N_{\text{train}} = 5000$, the rms in the $z < 0.5$ bin drops by about 20 per cent while the rms in the $z > 1.0$ bin drops by more than 50 per cent. Thus larger training sets are important for tying down the redshifts of rare objects, but if this is not of particular interest (e.g. many large-scale structure surveys) then one can manage with smaller training sets.

It is important to note that the size of N_{train} required to achieve a given accuracy depends on the variation inherent in the data set. In particular, for photometric redshifts it depends on the redshift range. In Sections 5–7, in which the same $H < 20.5$-SAM catalogue is used, the training set size will be fixed at 10 000. For low-redshift/shallow surveys (e.g. SDSS – see Section 8), smaller training sets should suffice.

5 COMPARISONS WITH THE TEMPLATE-FITTING METHOD

Fig. 6 compares the model redshifts of 2000 test galaxies with the redshifts estimated using a committee of five $6:6:6:6:1$ ANNs and a training set of 10 000 galaxies. As found in Section 4, the rms of $\Delta_z = z_{\text{model}} - z_{\text{phot}}$ is 0.12 over the redshift range $0 < z < 3.5$. Of course, the performance depends entirely on the particular set of filters and limiting magnitudes[2] so, for comparison, the results of the template-fitting code HYPERZ[3] (Bolzonella, Miralles & Pelló 2000) are also shown (using the eight synthetic Bruzual & Charlot 1993 GISSEL'98 evolving templates, which are distributed with HYPERZ, and dust extinction A_V in the range 0.0–1.2). The HYPERZ results are comparable but slightly worse. Formally, the rms of Δ_z is 0.26, but this relatively high value may be due more to the small number of complete misidentifications – against which the ANN seems to be more robust – than to a general increased scatter. However, as a further comparison, the scatter measured by the half-range of the

---

[2] The sample catalogue is of much poorer quality than the Hubble Deep Field, for example.
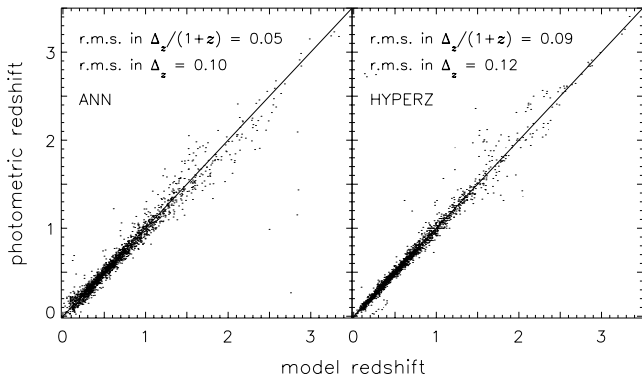
[3] http://webast.ast.obs-mip.fr/hyperz/

Figure 7. A comparison of ANN and HYPERZ photometric redshifts with model redshifts on 2000 test galaxies. The simulated catalogue is $H < 20.5$ with 5-σ limits $U = 25.1$, $B = 26.6$, $V = 26.1$, $R = 25.6$, $I = 24.7$ and $H = 20.5$. Prior to adding photometric noise, the photometry for each galaxy in the semi-analytic model catalogue was replaced with the photometry of the best-fitting of four empirical CWW template spectra (E, Sbc, Scd or Im) at the same redshift. A committee of five $6:6:6:1$ ANNs and a training set of 10 000 galaxies were used. The four CWW SEDs themselves were used in HYPERZ.

central 68 per cent Δ_z values is 0.02 for the ANN and 0.11 for HYPERZ.

It may be argued that HYPERZ is at a disadvantage here, relative to the ANN, since the ANN training and testing sets are based on the same spectral models while HYPERZ is trying to fit a different set of templates. For a real data set, HYPERZ – like all template-fitting methods – could still suffer from template mismatches while ANNs automatically fit the data. Synthetic templates were used in HYPERZ for the above comparison since synthetic templates, albeit from a different source, are used in the semi-analytic model. However when using HYPERZ on real data sets, empirical SEDs, such as the four CWW SEDs – E, Sbc, Scd and Im – distributed with the HYPERZ code, often produce better results. It is possible that the four CWW templates are a closer match to real galaxy SEDs than the eight GISSEL'98 evolving templates are to the semi-analytic model SEDs. Therefore as a further comparison, which maximally favours HYPERZ, the *UBV RIH* photometry for each galaxy in the semi-analytic model was replaced by *UBV RIH* photometry for the best-fitting (using rest-frame *B–I* colour) CWW template SED at the same model redshift. Noise was added to the new photometry as described above and an $H < 20.5$ sample was reselected. The magnitudes and redshifts were normalized as described above and a new committee of five $6:6:6:1$ ANNs was generated. Fig. 7 compares the new ANN with HYPERZ. The formal rms of Δ_z is 0.10 for the ANN and 0.12 for HYPERZ, while the half-width of the central 68 per cent Δ_z values is 0.02 for the ANN and 0.04 for HYPERZ. Hence, even for this case maximally favouring HYPERZ, the ANN performs at least as well as template fitting.

6 SCATTER DUE TO PHOTOMETRIC ERRORS

It is of interest to see what the distribution of estimated redshifts is for a particular galaxy, as a result of random photometric errors. Random noise was added, as described above, to a selection of individual model galaxies, with 1000 random simulations per galaxy. Then redshifts were estimated using the above previously-trained committee of five $6:6:6:1$ ANNs. Fig. 8 plots histograms of the estimated redshifts for each original galaxy. The width of the $P(z)$
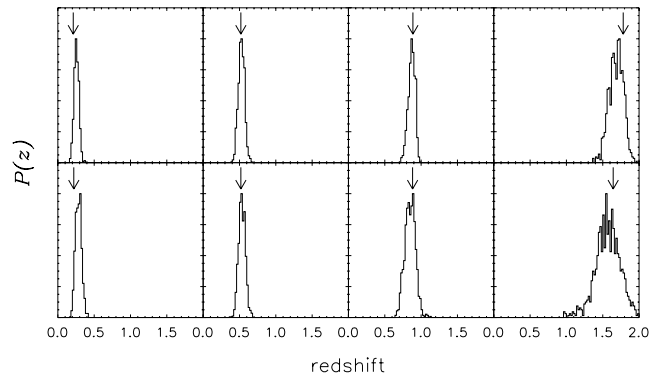

Figure 8. Eight galaxies were selected from the original 'noiseless' SAM catalogue. Random photometric noise was added simulating 5-σ limiting magnitudes $U = 25.1$, $B = 26.6$, $V = 26.1$, $R = 25.6$, $I = 24.7$ and $H = 20.5$, with 1000 simulations per initial galaxy. The plots show histograms (one for each of the original eight galaxies) of the redshifts estimated with the committee of five ANNs previously trained on the full 'noisy' $H < 20.5$ training set (as in Section 5). The upper four panels are for bright ($H \sim 19.5$) galaxies while the lower four panels are for faint ($H \sim 20.5$) galaxies. The arrows indicate the true (model) redshift of each galaxy.

distribution increases for fainter galaxies but a more pronounced trend is that it increases for galaxies at higher redshifts. There are fewer training galaxies at high redshifts which means that (a) the network weights are less constrained than at lower redshifts, and (b) the training process gives more weight to improving the accuracy for the majority of galaxies at low redshifts at the expense of accuracy at high redshifts.

7 SPECTRAL TYPE CLASSIFICATION

One can also use ANNs to determine spectral type (independently of redshift) provided, as emphasized above, the training set is representative in terms of both spectral types *and* redshifts (and hence galaxy colours). As in Section 5, the photometry for each semi-analytic model galaxy was replaced by that for the best-fitting CWW template SED at the same model redshift. Then simulated photometric noise was added and an $H < 20.5$ sample was selected. The best-fitting CWW spectral types E, Sbc, Scd and Im give the desired output on which the network is trained.

Since the input colour data is the same as in the photometric redshift problem of Section 4, a similar network architecture was used. However it was modified to produce four outputs (viz. $6:6:6:4$), corresponding to the four spectral types (*at any redshift*).[4] When training, the desired output is 1 for the output node corresponding to the correct type and 0 for the other three nodes. When a galaxy of unknown spectral type is run through the ANN, the output in each node may be treated approximately as a probability for the galaxy being of the corresponding type, and the galaxy is assigned the type for which the probability is greatest (cf. e.g. Storrie-Lombardi et al. 1992; Lahav et al. 1996). The input magnitudes were normalized to the range [0, 1] as above, 10 000 training objects were used, and a committee of five $6:6:6:4$ ANNs was generated.

---

[4] Alternatively, since galaxy spectral types roughly follow a sequence (e.g. Connolly et al. 1995a; Naim et al. 1995), one could utilize a network with a single output node to classify galaxies, for example, on a scale of 0 to 1, where 0 corresponds to spectral type E and 1 corresponds to spectral type Im.

Table 2. Comparisons of the efficiency with which ANNs and HYPERZ recover galaxy spectral types. The simulated catalogue is $H < 20.5$ with 5-σ limits $U = 25.1$, $B = 26.6$, $V = 26.1$, $R = 25.6$, $I = 24.7$ and $H = 20.5$. A committee of five 6:6:6:6:4 ANNs and a training set of 10 000 galaxies were used.

Estimated SED	HYPERZ				ANN			
	E	Sbc	Scd	Im	E	Sbc	Scd	Im
true SED								
E	428	1	0	0	422	6	0	1
Sbc	9	743	1	0	3	747	3	0
Scd	0	3	383	0	0	1	382	3
Im	0	0	9	423	3	1	7	421

The ANN results are displayed in Table 2. The ANN spectral types agree very well with the original CWW spectral types – the mean error rate is ∼1 per cent. Table 2 also shows the equivalent HYPERZ results. Here the ANN and HYPERZ perform comparably well.

8 PERFORMANCE OF NEURAL NETWORKS ON SDSS DATA

The Sloan Digital Sky Survey[5] (SDSS; York et al. 2000) consortium have now publicly released more than 50 000 spectroscopic redshifts along with *ugriz* photometry and various image morphological parameters. These provide an excellent opportunity to test ANNs on real data (see also Sowards-Emmerd et al. 2000 for a polynomial-fitting approach). Objects were selected from the SDSS public data set using the following criteria: (1) the spectroscopic redshift confidence must be greater than 0.95 and there must be no warning flags; (2) $r < 17.5$; (3) redshift <0.5. Stars were left in with the galaxies but at these magnitudes they could have been fairly robustly removed using image morphology. The order was randomized and the magnitudes, redshifts and other parameters were normalized to the range [0, 1].

Because the SDSS redshift range is much smaller than that of the SAM catalogue used in Section 4 (though note the addition of stars here), a simpler architecture may suffice. However, for the sake of simplicity, a similar architecture is used in this section also. Two ANN architectures were used. One – $5:6:6:6:1$ – inputing *ugriz* photometry, and the other – $8:6:6:6:1$ – inputing *ugriz* photometry and the SDSS pipeline star/galaxy classifier ('type') and Petrosian 50 per cent and 90 per cent *r*-band flux radii, r_{50} and r_{90}. A training set of size 10 000 was used and, for each architecture, a committee of five ANNs was generated.

Fig. 9 compares the ANN redshifts with spectroscopic redshifts for a testing set of 7000 galaxies. The rms's of Δ_z are 0.023 and 0.021 for committees of five $5:6:6:6:1$ and $8:6:6:6:1$ networks, respectively, while the mean offsets are both 0.000. These results are
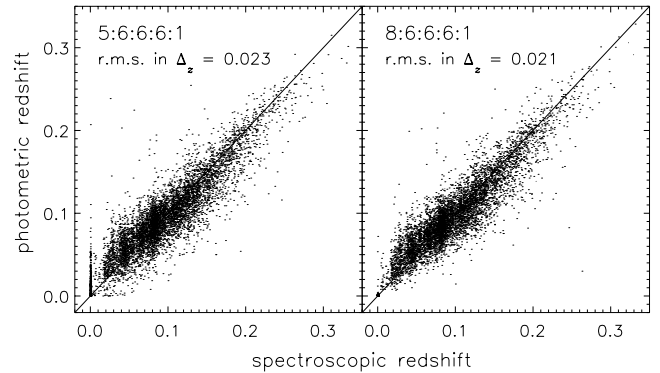
Figure 9. A comparison of photometric and spectroscopic redshifts using SDSS public data. Two ANN architectures were used, taking as input *ugriz* photometry (5:6:6:6:1 architecture, left) and *ugriz* photometry, SDSS star/galaxy classifier and Petrosian 50 and 90 per cent *r*-band flux radii (8:6:6:6:1 architecture, right). A training set of size 10 000 was used. The ANNs were tested on a separate testing set of size 7000 (plotted). In each panel, redshift estimates are medians from a committee of five ANNs.

easily as good as, and probably better than, the best results that template-fitting photometric redshift methods can produce. There are also very few outliers. The spike at $z = 0$ in the $5:6:6:6:1$ network results is due to misidentified stars. In the $8:6:6:6:1$ network, the addition of morphological parameters largely removes this feature.

While 10 000 training objects were used for the above networks, one can still do fairly well with much smaller training set sizes. For example, with only 500 training objects, the rms values become, respectively, 0.028 and 0.027 (cf. Fig. 5). Again, larger training sets are expected to be useful for pinning down the classifications of rare objects – e.g. the rms in the redshift range $0.25 < z < 0.35$ is ∼0.03 for $N_{\text{train}} = 10\,000$ but degrades to ∼0.06 for $N_{\text{train}} = 500$.

9 CONCLUSIONS

ANNs can produce photometric redshift accuracies that are comparable with, or better than, template-fitting procedures. However they do rely on large and representative training samples and an ANN is only applicable to the particular survey filters and redshift range upon which it has been trained. For large photometric/spectroscopic surveys, such as the SDSS and future deeper surveys such as DEEP2[6] and the VIRMOS-VLT Deep Survey[7] (VVDS; Le Fèvre et al. 2000), where large spectroscopic samples are available, it seems that ANNs offer some significant advantages over previous approaches. The VVDS, for example, is expected to include >100 000 redshifts to $I_{AB} = 22.5$, >40 000 redshifts to $I_{AB} = 24$ and >1000 redshifts to $I_{AB} = 25$, providing ample training set sizes for the complementary deep imaging in $UBVRIK_s$ to $I_{AB} = 25$ (similar to the limits used in Section 3).

With careful modelling of photometric errors and some loss in the Bayesian statistics, bright spectroscopic samples may also be extrapolated to provide training sets for fainter photometric samples.

In addition, ANNs may also be used where spectroscopic redshifts are unavailable, by utilizing a simulated catalogue (e.g. semi-analytic model) as a training set. By using theoretical SEDs in the training set, this method has all the disadvantages and advantages

of standard template fitting, but it also has the extra advantages (i) the 'template' SEDs include a (more or less) realistic distribution of *complex* star formation histories, dust modelling and metallicities, etc., giving fully Bayesian statistics, and (ii) the weights applied to different filters (and non-linear combinations thereof) may be more optimal than simple χ^2-weighting.

To conclude, while template-fitting photometric redshifts may be used to good effect in pioneering studies of new populations of objects, spectroscopic confirmation will always be necessary to obtain truly robust scientific results. Instead, the real power of photometric redshifts lies in extending small very resource-intensive faint spectroscopic surveys to much larger fields of view and sample sizes. That is to say, the area where photometric redshifts can best be used for robust and useful scientific gain is in the training-set regime. ANNs provide a powerful tool for obtaining high-quality photometric redshifts in such surveys.

REFERENCES

Allende Prieto C., Rebolo R., López R. J. G., Serra-Ricart M., Beers T. C., Rossi S., Bonifacio P., Molaro P., 2000, AJ, 120, 1516

Andreon S., Gargiulo G., Longo G., Tagliaferri R., Capuano N., 2000, MNRAS, 319, 700

Bailer-Jones C. A. L., Irwin M., von Hippel T., 1998, MNRAS, 298, 361

Baum W. A., 1962, IAU Symp. No. 15, Problems of Extragalactic Research, p. 390

Benitez N., 1998, in Paul J., Montmerle T., Aubourg E., eds, Abstracts of the 19th Texas Symposium on Relativistic Astrophysics and Cosmology

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bishop C. M., 1995, Neural Networks for Pattern Recognition. Oxford Univ. Press

Bolzonella M., Miralles J.-M., Pelló R., 2000, A&A, 363, 476

Brunner R. J., Szalay A. S., Connolly A. J., 2000, ApJ, 541, 527

Bruzual A. G., Charlot S., 1993, ApJ, 405, 538

Cole S., Aragón-Salamanca A., Frenk C., Navarro J., Zepf S., 1994, MNRAS, 271, 781

Coleman G. D., Wu C.-C., Weedman D. W., 1980, ApJS, 43, 393 (CWW)

Connolly A. J., Szalay A. S., Bershady M. A., Kinney A. L., Calzetti D., 1995a, AJ, 110, 1071

Connolly A. J., Csabai I., Szalay A. S., Koo D. C., Kron R. G., Munn J. A., 1995b, AJ, 110, 2655

Connolly A. J., Szalay A. S., Brunner R. J., 1998, ApJ, 499, L125

Devriendt J. E. G., Guiderdoni B., Sadat R., 1999, A&A, 350, 381

Fernández-Soto A., Lanzetta K. M., Yahil A., 1999, ApJ, 513, 34

Fernández-Soto A., Lanzetta K. M., Chen H.-W., Pascarelle S. M., Yahata N., 2001, ApJS, 135, 41

Firth A. E. et al., 2002a, MNRAS, 332, 617

Firth A. E., 2002b, PhD thesis., Univ. Cambridge

Fontana A., D'Odorico S., Poli F., Giallongo E., Arnouts S., Cristiani S., Moorwood A., Saracco P., 2000, AJ, 120, 2206

Gwyn S. D. J., Hartwick F. D. A., 1996, ApJ, 468, L77

Hogg D. W. et al., 1998, AJ, 115, 1418

Kauffmann G., White S., Guiderdoni B., 1993, MNRAS, 264, 201

Lahav O., Naim A., Sodré L., Jr, Storrie-Lombardi M. C., 1996, MNRAS, 283, 207

Le Fèvre et al., 2000, in Clowes R. G., Adamson A. J., Bromage G. E., eds, ASP Conf. Ser. Vol. 232, New Era of Wide-Field Astronomy. Astron. Soc. Pac., San Francisco, p. 449

Massarotti M., Iovino A., Buzzoni A., 2001a, A&A, 368, 74

Massarotti M., Iovino A., Buzzoni A., Valls-Gabaud D., 2001b, A&A, 380, 425

Naim A., Lahav O., Sodre L., Jr, Storrie-Lombardi M. C., 1995, MNRAS, 275, 567

Sheth R. K., Tormen G., 1999, MNRAS, 308, 119

Somerville R. S., 1997, PhD thesis., Univ. California

Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504

Sowards-Emmerd D., Smith J. A., McKay T. A., Sheldon E., Tucker D. L., Castander F. J., 2000, AJ, 119, 2598

Storrie-Lombardi M. C., Lahav O., Sodré L., Jr, Storrie-Lombardi L. J., 1992, MNRAS, 259, 8

Teplitz H. I., Hill R. S., Malumuth E. M., Collins N. R., Gardner J. P., Palunas P., Woodgate B. E., 2001, ApJ, 548, 127

Weaver W. B., 2000, ApJ, 541, 298

Weymann R. J., Storrie-Lombardi L., Sawicki M., Brunner R. J., eds, 1999, ASP Conf. Ser. 191, Photometric Redshifts and High Redshift Galaxies. Astron. Soc. Pac., San Francisco

York D. G. et al., 2000, AJ, 120, 1579

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