Summer 7-1997

Physical Parameterization of Stellar Spectra: The Neural Network Approach

Coryn A.L. Bailer-Jones
Institute of Astronomy

Ted von Hippel
University of Wisconsin, vonhippt@erau.edu

Mike Irwin
Department of Astronomy, University of Wisconsin

Gerard Gilmore
Institute of Astronomy, gil@ast.cam.ac.uk

Mon. Not. R. Astron. Soc. **000**, 000–000 (1997) Printed 12 November 2018 (MN LATEX style file v1.4)

Physical Parameterization of Stellar Spectra: The Neural Network Approach

Coryn A.L. Bailer-Jones[1⋆†], Mike Irwin[2], Gerard Gilmore[1] and Ted von Hippel[3]

[1] *Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA, UK*
[2] *Royal Greenwich Observatory, Madingley Road, Cambridge, CB3 0EZ, UK*
[3] *Department of Astronomy, University of Wisconsin, Madison, WI 53706, USA*

Submitted April 1997, Accepted July 1997

ABSTRACT

We present a technique which employs artificial neural networks to produce physical parameters for stellar spectra. A neural network is trained on a set of synthetic optical stellar spectra to give physical parameters (e.g. $T_{\rm eff}$, $\log g$, [M/H]). The network is then used to produce physical parameters for real, observed spectra.

Our neural networks are trained on a set of 155 synthetic spectra, generated using the SPECTRUM program written by Gray (Gray & Corbally 1994, Gray & Arlt 1996). Once trained, the neural network is used to yield $T_{\rm eff}$ for over 5000 B–K spectra extracted from a set of photographic objective prism plates (Bailer-Jones, Irwin & von Hippel 1997a). Using the MK classifications for these spectra assigned by Houk (1975, 1978, 1982, 1988), we have produced a temperature calibration of the MK system based on this set of 5000 spectra. It is demonstrated through the metallicity dependence of the derived temperature calibration that the neural networks are sensitive to the metallicity signature in the real spectra. With further work it is likely that neural networks will be able to yield reliable metallicity measurements for stellar spectra.

Key words: methods: data analysis, numerical - stars: fundamental parameters

1 INTRODUCTION

The MK classification system was first proposed in its current form in 1943 by Morgan, Keenan & Kellman (1943), and has since undergone a number of revisions (e.g. Keenan & McNeil (1976), Morgan, Abt & Tapscott (1978)). MK classification is the only widely used system for stellar spectral classification. Over its history it has contributed towards a number of important developments in astronomy, such as the further development of the now-famous HR diagram (Hertzprung 1911, Russell 1914) and the identification of anomalous stars. Currently, MK classification is largely used as a tool in the preliminary analysis of unusual stars, and in selecting stellar samples for further study.

An often-stated advantage of the MK system is that its classifications, often based upon the visual inspection of spectra, are static because they are based on a set of standards. However, a given spectrum may be classified differently by different people, and any one person may also classify a given spectrum differently at different times. These problems of subjectivity could be partially alleviated through the use of automated classifiers (von Hippel et al. 1994, Bailer-Jones et al. 1997a). Automated classifiers could also produce quantitative errors associated with their classifications. Another problem with the MK system is that it lacks a well-defined metallicity parameter, whereas metallicity variations are known to have a significant effect on the appearance of high (~ 1 Å) resolution spectra. This limits the system to classifications of bright, nearby stars which show only limited metallicity variations.

Attempts to extend and revise MK classification (e.g. Corbally, Gray & Garrison 1994) may well prove valuable, but as our understanding of stellar spectra grows, particularly from computational work with model atmospheres and synthetic spectra, it becomes increasingly desirable to obtain reliable physical parameterizations of stars. Any classification system is a compromise between retaining the full information in the spectrum and the need for a compact summary of it. The optimal 'summary' is of course given by the physical parameters. Advances in computational power and data storage since the inception of MK classification mean that it is now practicable to process and store large numbers of spectra. The development of fast, automated

⋆ Present Address: Mullard Radio Astronomy Observatories, Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE, UK
† email: calj@mrao.cam.ac.uk

classifiers will mean that it is feasible to 'classify' large numbers of stellar spectra in terms of their physical parameters and to re-classify them rapidly whenever stellar models are improved. Physical parameters should be obtained from an original spectrum, rather than an empirical classification, as the latter may well disregard certain spectral features which later turn out to be important. One of the advantages of the MK classification system is that it is an empirical system based on unchanging standards, whereas any direct parameterization of a spectrum depends upon the quality of stellar models and will change as these improve. MK classifications could remain as useful labels giving a rough 'feel' for a spectrum, and the work of von Hippel et al. (1994) and Bailer-Jones et al. (1997a) has shown that reliable automated MK classification is possible. In this paper we demonstrate how we extend our automated techniques to the determination of physical parameters directly from an observed spectrum.

2 OBSERVATIONAL DATA

The observational stellar spectra used in this project were taken from objective prism plates obtained in the Michigan Spectral Survey (Houk 1994). This was an objective prism survey of the whole southern sky ($\delta < +12^\circ$) from the Curtis Schmidt telescope at the Cerro Tololo Interamerican Observatory in Chile. We were loaned 100 of these plates by Houk, which we digitized using the APM facility in Cambridge. The output from the APM is an optical density. Cawson et al. (1987) showed that the optical density calculated by the APM is, to a good approximation, linearly related to the incident intensity on the photographic plate for the typical range of optical density encountered for this type of extraction. Once scanned, the digitized plates were reduced and the spectra extracted off-line. This yielded a set of over 5000 spectra over the approximate spectral range 3800–5200 Å at a resolution of ∼3 Å. This data set covers a wide range of spectral types (B2–M7) for luminosity classes III, IV and V. The Michigan Spectral Survey was designed to be a reclassification of the Henry Draper stars. In keeping with previous usage, we shall refer to our set of 5000 spectra from this survey as the MHD (Michigan Henry Draper) spectra. Further details of this data set and the spectral extraction technique are discussed in Bailer-Jones, Irwin & von Hippel (1997b). These spectra have been used in a related project to automate MK stellar classification using a neural network (Bailer-Jones et al. 1997a).

3 NEURAL NETWORKS

A neural network is a software device which can be trained to give a non-linear parameterized mapping between a number of inputs (e.g. a complete stellar spectrum) and one or more outputs (e.g. parameters such as $T_{\rm eff}$ or MK spectral type). Figure 1 shows a neural network architecture with a single output. Each node in the input layer holds a value, I_i. In our application the vector of inputs, $(I_1, I_2, \ldots, I_i, \ldots)$, is a stellar spectrum and the output is the effective temperature, $T_{\rm eff}$. Each of the input nodes connects to every node in the next layer of nodes, the 'hidden' layer, and each of these connections has a weight, $w_{i,j}$, associated with it. The j^{th} node in the hidden layer forms a weighted sum of its inputs, given by $x_j = \sum_i w_{i,j} I_i$. It then passes this sum through a non-linear sigmoid transfer function to give the final output from this node, $H_j = (1 + e^{-\lambda x_j})^{-1}$, where λ is some constant. These outputs from the nodes in the hidden layer then serve as the inputs to the node in the output layer, which performs the same processing. The output from this final node is the network output. The non-linearity of the sigmoid function means that the network output is a non-linear function of the inputs. It can be shown that neural networks can model functions of arbitrary complexity (e.g. Bishop (1995) and references therein). Neural networks are therefore useful in a wide range of data modelling applications.

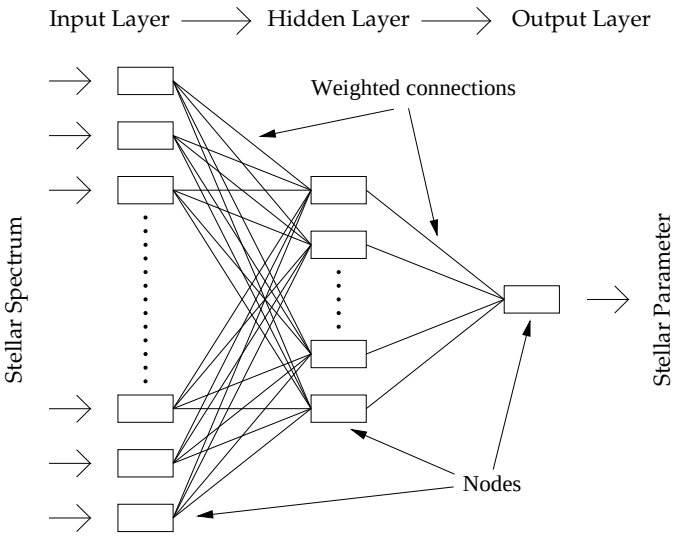

Figure 1. Neural network architecture. A neural network consists of layers of nodes linked by weighted connections (weights). Each node in the hidden and output layers performs a non-linear weighted sum of its inputs which it then passes to its output. Therefore the final output from the neural network is a non-linear function of the network inputs parameterized by the weights. The weights are found by 'training' the neural network using a set of inputs for which the ideal outputs (the 'target outputs') are known. This training process is therefore equivalent to interpolating the training data to find the underlying function relating the inputs to the output.

The key to producing the desired output from the network when it is presented with a certain input is to set the network parameters – the weights – to their correct values. This is known as 'training' the network and requires a set of input data for which the associated ideal outputs (the 'target outputs') are known. Training takes place as follows. The weights are initially set with random values over a small range. Thus when a spectrum is fed into the network, the output will also be random. By comparing this output with what it should be (the target output), we can adjust the weights to give an output which is closer to the target value. This is repeated for each input/output pair in the training data set. The network is trained iteratively by successive passes of the training data through the network, and on each pass the weights are perturbed towards their optimal values. Specifically, the network training is performed by minimizing the error function $E = \sum_k (O_k - T_k)^2$, where

O_k is the network output for a given input spectrum and T_k is the target output for that spectrum. Thus we can think of training a neural network as an N-dimensional minimization problem in which we want to find those values of the N network weights which minimize E. Once the weights have been found they are fixed and the neural network can be applied to any number of new inputs for which the outputs are not known. Training the neural network is simply the process of interpolating the multi-dimensional training data in order to produce an input-output mapping characteristic of the problem represented by these data.

The output from a neural network is some non-linear function of all of the network inputs. In our particular application this means that the values of physical stellar parameters which the neural network gives at its output are based on the appearance of the whole spectrum: we do not have to tell the network in advance which spectral lines are relevant. Based only upon the training data, the neural network will learn which wavelengths are more significant than others in determining the correct spectral parameters, and will express this by assigning appropriate values to the network weights.

Neural networks have been applied in a number of areas of astronomy, as reviewed by Storrie-Lombardi & Lahav (1994) and Miller (1993). Further theoretical details of neural networks can be found in, for example, Hertz, Krogh & Palmer (1991), Bishop (1995), Lahav et al. (1996) and Bailer-Jones (1996).

4 CALIBRATION PROCEDURE

In the following sections, theoretical stellar spectra of given physical parameters (such as $T_{\rm eff}$ and $\log g$) are generated and processed to the same 'flux' and dispersion scales as the observed MHD spectra described in the previous section. A neural network is trained on these synthetic spectra with their physical parameters as the network target outputs. Once trained, the neural network is used to give physical parameters for the MHD stars. This provides a means of rapidly and easily obtaining physical parameters for a large number of spectra without calibration via the MK system. However, a statistical comparison of these physical parameters with the known MK classifications of the MHD stars provides a calibration between the MK system and physical stellar parameters. Training the network is very quick (~ 10 seconds on a SUN Sparc 10), and applying it is even faster, so it is not prohibitive to re-train the network as improved models are obtained.

5 SYNTHETIC SPECTRA

5.1 Generation

The synthetic spectra were created using the SPECTRUM program written by Gray (Gray & Corbally 1994, Gray & Arlt 1996). This program computes a synthetic spectrum under the assumption of Local Thermodynamic Equilibrium (LTE) using a model stellar atmosphere. We used the fully blanketed models calculated by Kurucz (1979, 1992). The model atmosphere is a tabulation of temperature and pressure at a range of 'mass depths' in the stellar photosphere, calculated on the basis of a variety of sources of opacity. These model atmospheres are characterized by four parameters: metallicity, [M/H]; microturbulence velocity, $V_{\rm micro}$; surface gravity, $\log g$; effective temperature, $T_{\rm eff}$. Each spectrum is uniquely labelled by these four parameters. To calculate the spectrum emergent from the model atmosphere, SPECTRUM also requires a table of atomic and molecular species which lists their relative abundances, masses, and ionization energies (or disassociation energies for molecules). SPECTRUM is also given a line list, listing, for each line in the spectrum, the atomic or ionic species producing the line, the energy of the two bound electron energy levels involved in the transition, the oscillator strength of the transition and a damping factor. From these, SPECTRUM calculates the densities of electrons, atoms and ions at different layers in the atmosphere, from which it determines the atmospheric opacity as a function of optical depth and wavelength using a number of different opacity sources (Rayleigh scattering, electron scattering, bound-free opacities etc.). Under the assumption of LTE, the source function, S_ν, is equal to the Plank function, and the latter can be computed directly from the temperatures tabulated in the model atmosphere. Using the source function and opacities, SPECTRUM calculates the synthetic spectrum in small wavelength portions. This is done by calculating the continuum and line absorption strengths at each wavelength and evaluating a Voigt broadening profile for the lines using van der Waals, natural, and quadratic Stark broadening. More details of SPECTRUM can be found in Gray & Corbally (1994).

Table 1 lists the spectra which we generated. The $\log g = 4.0$ and $\log g = 4.5$ models are dwarfs and the remainder correspond to giants or subgiants. The spectra in Table 1 were calculated for each of four metallicities, [M/H] $= 0.0, -0.2, -0.5, -1.0$, giving a total of 155 spectra at each metallicity. A microturbulence velocity of $V_{\rm micro} = 2.0\,{\rm km\,s^{-1}}$ was used throughout. SPECTRUM creates good quality spectra over a fairly wide range of spectral types, but it does not produce very reliable spectra for very early- or very late-type stars. The generation of accurate spectra at the hot end is inhibited primarily by the assumption of LTE, as NLTE effects are important in O-type stars. This should not have any significant effect on the calibrations of the MHD spectra, as the earliest type stars present are B2. At the other end of the temperature scale, spectra cooler than about 4250 K (approximately K7 and later for dwarfs, and K3 and later for giants) are not accurately synthesized on account of the absence of many important molecules (e.g. H_2O) from Kurucz's models, and on account of the fact that SPECTRUM does not include TiO in its opacity calculations (although it does include MgH, C_2, NH, CH, CN and SiH) (R. Gray, private communication, 1996). The quality of the spectra will also start to decrease for stars later than about G2 (solar), because in these stars a significant part of the line formation occurs in the chromosphere, which is not modelled by SPECTRUM.

5.2 Processing

The synthetic spectrum generated by SPECTRUM is an energy flux spectrum evaluated at 0.02 Å intervals. If a neural network is to be trained on a set of synthetic spectra and used to get physical parameters for real observed spectra, then

Table 1. Synthetic spectra generated and used for network training. A spectrum was generated at each of the four metallicities, [M/H] = 0.0, −0.2, −0.5 and −1.0, yielding a set of 155 spectra per metallicity. The microturbulence velocity, $V_{\rm micro}$, was fixed at 2.0 km s^{-1}. The temperature steps above are the finest steps in which Kurucz's model atmospheres are available. The gaps at high temperature represent unstable model atmospheres which blow apart due to high radiative fluxes (Kurucz 1992). SPECTRUM takes about one hour (on a SUN Sparc 10) to calculate a synthetic spectrum in 0.02 Å steps over the range 3790 Å–5200 Å.

$T_{\rm eff}$	log g					
	2.0	2.5	3.0	3.5	4.0	4.5
30000				x		
25000			x	x		
22000					x	x
20000			x	x	x	x
18000					x	x
16000					x	x
15000		x	x	x		
14000	x	x	x	x	x	x
13000		x	x	x	x	x
12000		x	x	x	x	x
11000		x	x	x	x	x
10000	x	x	x	x	x	x
9750					x	x
9500	x	x	x	x	x	x
9250					x	x
9000	x	x	x	x	x	x
8750					x	x
8500	x	x	x	x	x	x
8250					x	x
8000	x	x	x	x	x	x
7750					x	x
7500	x	x	x	x	x	x
7250					x	x
7000	x	x	x	x	x	x
6750					x	x
6500	x	x	x	x	x	x
6250					x	x
6000	x	x	x	x	x	x
5750	x	x	x	x	x	x
5500	x	x	x	x	x	x
5250	x	x	x	x	x	x
5000	x	x	x	x	x	x
4750	x	x	x	x	x	x
4500	x	x	x	x	x	x
4250	x	x	x	x	x	x
4000	x	x	x	x	x	x

the observed and synthetic spectra must be processed into a homogenous form. Specifically, they must have common wavelength and flux scales. Because we are free to generate the synthetic spectra at high resolution and infinite S/N, we chose to process these into the format of the MHD spectra, rather than vice versa. The processing steps are summarized in Figure 2.

When non-linear processing operations are involved, the order of the operations in this processing is relevant. It should correspond to the effective processing performed by the telescope, disperser and detector in obtaining a real spectrum. The first stage of processing the synthetic spectra was therefore to re-bin them (conserving flux in the process) to

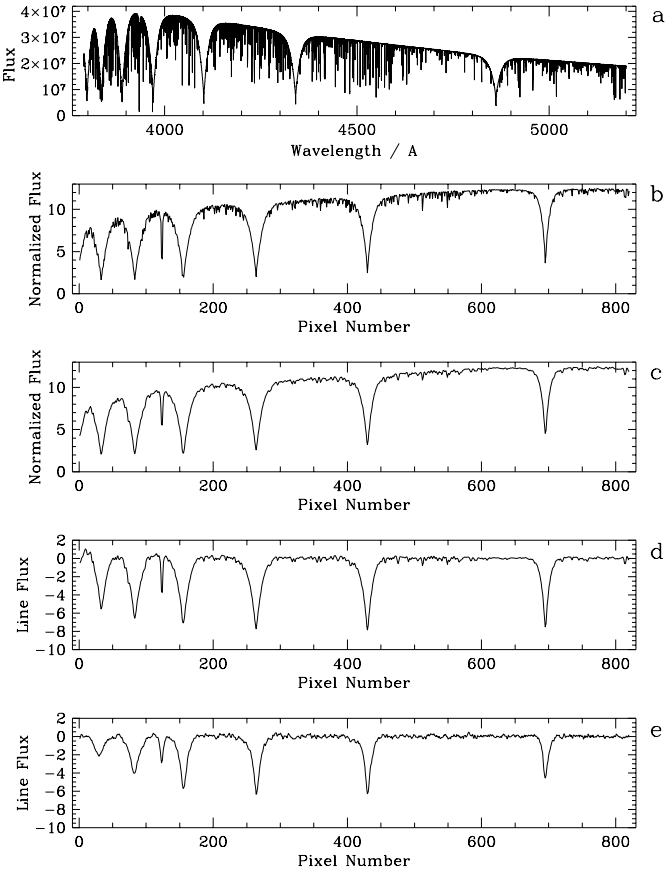

Figure 2. Processing a synthetic spectrum. (a) A synthetic spectrum as calculated by SPECTRUM, with parameters [M/H] = 0.0, log g = 4.0, $T_{\rm eff}$ = 9000, shown sampled at 0.1 Å on a linear wavelength scale. The flux is in units of $erg\,s^{-1}\,cm^{-2}\,\AA^{-1}$. (b) The spectrum is re-binned to the dispersion of the MHD spectra and area normalized. (c) The spectrum is blurred with a $\sigma = 0.88\,pix$ (= 3″ FWHM 'seeing') Gaussian filter. (d) The continuum is removed. (e) A bright (mag = 8.1) MHD spectrum (HD 23164) of spectral class A1 V, which has similar physical parameters to the synthetic spectrum shown in (d).

the dispersion of the MHD spectra. This required the wavelength calibration of the MHD spectra. We obtained this from the average spectrum of the MHD data set to give a calibration which averaged out any slight wavelength misalignments incurred when reducing the objective prism spectra. A third-order fit to 16 identified lines in the spectrum gave an RMS wavelength–pixel calibration error of < 0.4 Å. As the synthetic spectra are over-sampled relative to the real spectra by a factor of at least 50 (= 1.05 Å pix^{-1}/0.02 Å), we were justified in linearly interpolating the synthetic spectra across their 0.02 Å-wide bins when re-binning them to the MHD flux bins. This re-binning yields a spectrum such as that shown in Figure 2b. Note that the apparent number and strength of the metal lines below some equivalent width is greatly reduced.

The next step was to 'blur' the spectra to simulate atmospheric turbulence ('seeing') and telescope tracking errors. This blurring was achieved by convolving a Gaussian (truncated at $\pm 3\sigma$) with the spectrum. This Gaussian was normalized to conserve flux. A smooth blurring (and flux

conservation) at the end of the spectrum was obtained by reflecting the spectrum about its ends prior to the convolution. We experimented with a number of different values of the σ parameter to obtain a best match between the appearance of a blurred synthetic spectrum and an MHD spectrum of similar physical properties (as inferred from the MK classification and visual inspection). Given that each bin in an MHD spectrum corresponds to 1.45″, a FWHM (2.35σ) seeing of θ arcseconds corresponds to $\sigma = 0.29 \times \theta\, pix^{-1}$ arcseconds. We found that the optimal blurring was about 3″ (0.88 pixels), which seems reasonable given the relatively poor tracking ability of the Michigan Curtis Schmidt telescope.

The final stage of preprocessing was continuum removal. This was done in exactly the same way as for the MHD spectra: a spectral continuum is produced by median and then boxcar filtering the spectrum. This continuum is subtracted from the original spectrum to yield a rectified (line-only) spectrum. To improve the continuum fit in the region of broad lines (most importantly H lines, the CN band and Ca II H&K lines), these regions were masked off prior to the median filtering. Further details are given in Bailer-Jones et al. (1997b).

The final spectrum is shown in Figure 2d above a real spectrum of similar physical parameters. The removal of the continuum from the synthetic spectra gives a good line-only spectrum everywhere apart from at the very blue end. This was not apparent in the continuum removal of the MHD spectra as the plate QE is very low at the blue end of the spectrum. To improve the match between the spectra, we chose to exclude the first ten flux bins from all synthetic and MHD spectra when using them together in a neural network. The final spectra therefore consist of 810 flux bins.

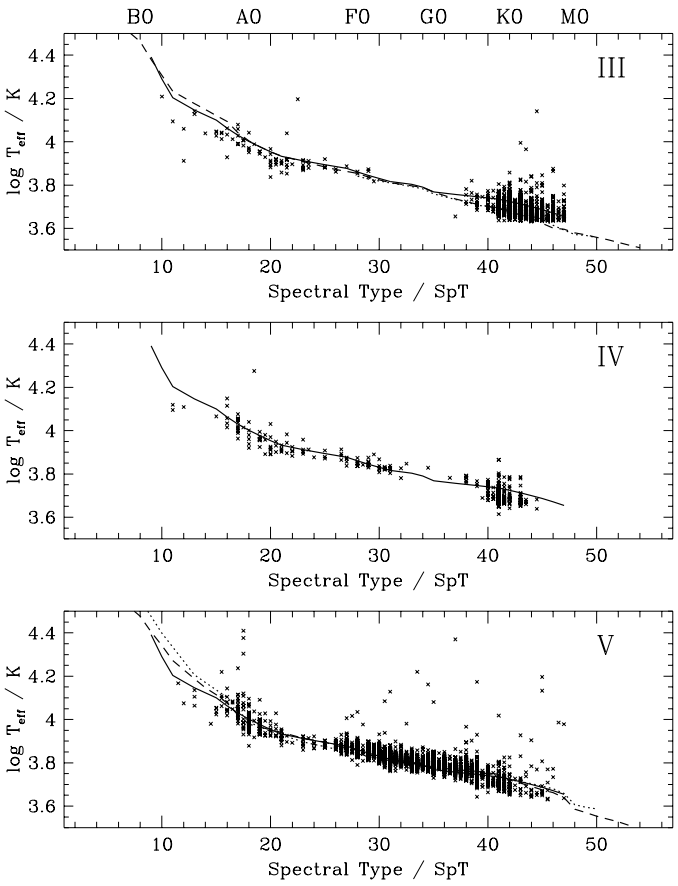

Figure 3. $T_{\rm eff}$–SpT relationship for the MHD spectra from a committee of neural networks trained on [M/H] = −0.2 synthetic spectra. The relationships for luminosity classes III, IV and V are shown separately, although the values of $T_{\rm eff}$ for these were obtained from the same committee. The lines show calibrations from the literature for comparison. The solid line is the calibration of Gray and Corbally (1994); the dashed line is that of Schmidt-Kaler (1982); the dotted line is that of Gray (1992). Gray and Corbally only give a calibration of dwarf (V) stars, but this calibration line has been included in the top two plots to ease comparison between the plots of the neural network calibrations. Schmidt-Kaler and Gray both give separate calibrations for giants (III). None give calibrations for the subgiants (IV).

6 NEURAL NETWORK CALIBRATION

6.1 $T_{\rm eff}$ Calibration

To determine effective temperature, $T_{\rm eff}$, for the MHD spectra, a neural network was set-up with a single continuous output to represent $T_{\rm eff}$. The type of neural network used had an output range confined to 0–1. In order to fit the 4000 K to 30 000 K temperature range of the synthetic models comfortably into this range, we used the transformation $T = (\log T_{\rm eff} - 3)/2$, where T is the network target. A more straightforward implementation would be to use a network with linear outputs which has an unbounded output range, although we would probably still want to use $\log T_{\rm eff}$ as the target in order to reduce the dynamic range. The exact transformation used should not matter.

Following on from the work of Bailer-Jones et al. (1997a), we used an 810:5:5:1 neural network architecture for this calibration problem. This refers to a neural network with 810 inputs (the number of flux values in the spectrum), 5 nodes in each of two hidden layers (Figure 1 shows a network with only a single hidden layer) and a single output.

Rather than using a single neural network to give predictions of $T_{\rm eff}$, we used a committee of ten neural networks differing only in their initial random weights. The ouputs given by each member of the committee are averaged to give a single prediction which is generally more reliable than that given by any single network. Four training data sets were constructed, one for each of the four metallicities at which synthetic spectra were generated: [M/H] = 0.0, −0.2, −0.5 and −1.0. Thus each training set consisted of 155 spectra of the same metallicity and with values of $T_{\rm eff}$ and $\log g$ shown in Table 1. A separate committee was trained on each of the four data sets. Once trained, the committees were used to evaluate $T_{\rm eff}$ for all of the MHD spectra. Figure 3 shows the $T_{\rm eff}$-SpT calibration produced by the committee trained on the [M/H] = −0.2 spectra, over-plotted by calibrations from the literature.

The MK spectral type classifications along the x-axis are the catalogue classifications of the MHD spectra as given by Houk (1975, 1978, 1982, 1988), converted into codes on a 1–57 numerical scale as shown in Table 2. These classifications are all either integral or half-integral spectral subtypes, so the data are discrete along the x-axis. Because the syn-

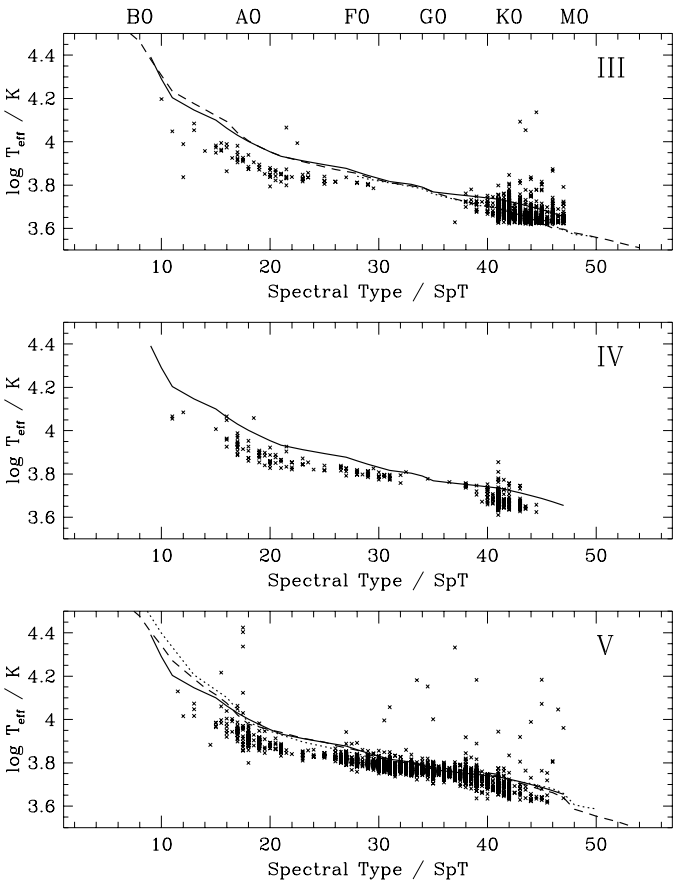

Figure 4. $T_{\rm eff}$–SpT relationship for the MHD spectra from a committee of neural networks trained on [M/H] = −1.0 synthetic spectra. See caption to Figure 3 for further details.

thetic spectra are only valid for temperatures above about 4250 K, we would not expect the neural networks to give reliable calibrations for spectra later than about K5, and we have omitted all spectra later than K5 from these calibration plots.

Apart from a handful of outliers, the majority of the 2732 dwarfs in the bottom panel of Figure 3 show a reasonably 'tight' correlation between MK spectral type and $T_{\rm eff}$. Their distribution also agrees quite well with the published calibrations. The giant stars also show a reasonable $T_{\rm eff}$–SpT correlation, although it deteriorates towards cooler stars, most probably due to the lower quality and accuracy of the synthetic spectra in this region. We have not discovered any $T_{\rm eff}$–SpT calibrations in the literature with which to compare the subgiant calibrations, but our data shows a reasonably tight correlation between $T_{\rm eff}$ and spectral type.

As a comparison, Figure 4 shows a similar plot to Figure 3 for a committee of ten networks trained on the [M/H] = −1.0 spectra. Here we see a slightly tighter correlation for the dwarfs, but the agreement with the published calibrations is not as good. The correlations for networks trained on [M/H] = −0.5 and [M/H] = 0.0 spectra were also quite tight, but they did not agree with the literature calibrations for both giants and dwarfs as well as [M/H] = −0.2. Metallicity effects on the calibration are discussed in section 6.2 below.

The data in Figure 3 can be used to give a statistical calibration between $T_{\rm eff}$ and spectral type. The most suitable way of achieving this is by forming the frequency distribution, $D_s(T_{\rm eff})$, of $T_{\rm eff}$ for each spectral type separately, where s indicates that D is different for each spectral type. With a sufficiently large amount of data, $D_s(T_{\rm eff})$ would be approximately Gaussian. The median of $D_s(T_{\rm eff})$, $\tilde{T}_{\rm eff}$, is a robust measure of the $T_{\rm eff}$ calibration for this spectral type. An appropriate measure of the spread of $D_s(T_{\rm eff})$ is σ_{68}. This is the value which confines 68% of the data, and is equal to 1σ if the distribution is Gaussian. For some spectral types there were only a few spectra; in these cases it was necessary to linearly interpolate $D_s(T_{\rm eff})$ in order to yield an accurate measure of σ_{68}.

Figure 5 plots these calibrations and their σ_{68} errors. It shows a very good agreement with the published calibrations, particularly for the dwarf stars. At the ends of the spectral type scale the calculated errors become significantly larger. This could be partly on account of the increased difficulty of the neural network to interpolate its mapping function at the ends of the output range, as the interpolation is less well constrained in these regions. The larger errors at the cool end are more likely, however, to be attributable to the lower quality of the stellar models and synthetic spectra at low temperatures. The larger errors at high temperatures, SpT $\lesssim 20$, are probably an indication of a poorer determination of the $T_{\rm eff}$–SpT relationship which in turn is the result of there being fewer spectra in this region. Table 2 tabulates the calibrations and errors in Figure 5 and shows the number of spectra, N, of each spectral type used to determine $\tilde{T}_{\rm eff}$ and σ_{68}. Note that the literature calibrations will also have some degree of uncertainty, although as the respective authors do not provide details of this it is difficult to perform rigorous statistical tests of the level of agreement between our calibrations and the literature ones.

It is important to realise that the MK spectral type parameter is not a continuous variable. It is a set of discrete classes, each of which does not correspond to a single unique temperature. There is, therefore, a cosmic scatter of temperature about the median calibration value, $\tilde{T}_{\rm eff}$, and this intrinsic scatter will make a significant contribution to σ_{68} (other contributions are discussed below). The important point is that σ_{68} is *not* a measure of the error with which $\tilde{T}_{\rm eff}$ has been determined. The error in $\tilde{T}_{\rm eff}$ itself is given by the standard error in the median, which is smaller than σ_{68} by a factor of $0.79\sqrt{N}$ ($\equiv \sqrt{2N/\pi}$) for large N. This error for F5 dwarfs, for example, is $280/(0.79\sqrt{269}) = 22$ K. Almost all of the values of $D_s(T_{\rm eff})$ in Table 2 have standard errors in the median of less than 50 K. This is about as small as can be meaningfully reported. In comparison, Jones, Gimore & Wyse (1996) used a broad band photometric index, $(V-I)_c$, to determine temperature. They showed that an error in this index of $\pm 0\overset{m}{.}05$ gives rise to a temperature error of 120–240 K (depending upon the spectral type). An error of $\pm 0\overset{m}{.}02$ (which is about the limit of the photometry) would yield errors of 47–97 K.

Cosmic scatter is not the only factor contributing to σ_{68}. Another source of error is the simplifying assumptions used in generating the atmospheric models and synthetic spectra, notably the assumption of LTE and the absence of certain atoms and molecules. Additionally, the synthetic spectra do not display variances in other stellar properties which will

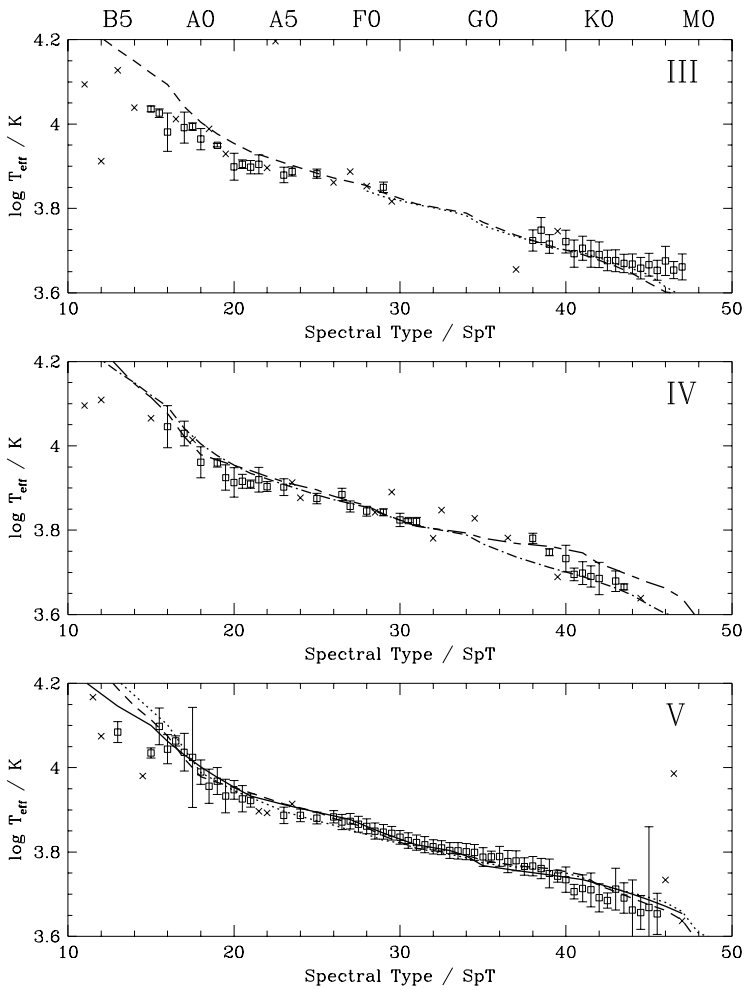

Figure 5. $T_{\rm eff}$–SpT calibration of data shown in Figure 3 ([M/H] = −0.2). The squares correspond to the median, $\tilde{T}_{\rm eff}$, of the frequency distribution of $T_{\rm eff}$ for each spectral type. The error bars are the σ_{68} errors, i.e. those bounds which include 68% of the data. Those spectral types for which there were only one or two points in the distribution are marked with a cross. These points are not very reliable and as the error cannot be reliably determined, it is not shown. The overplotted lines are the calibrations of Gray and Corbally (1994) (solid line), Schmidt-Kaler (1982)(dashed line) and Gray (1992) (dotted line). We found no published calibration for the subgiants (IV) and instead have overplotted the middle figure with the giant (dot-dash line) and dwarf (dash line) calibrations of Schmidt-Kaler. Note that the axis scales are slightly different here from those used in Figure 3.

show cosmic variance, such as the relative abundances of atomic and molecular species or $V_{\rm micro}$. These will produce variations in the MHD spectra which are 'unexpected' by a network trained on the synthetic spectra. Another source of error is likely to derive from the neglect of line formation mechanisms in the chromosphere.

Whilst we made efforts to process the synthetic spectra into the same format as the MHD spectra, they still differ in the possibly important respect that the synthetic spectra are on a linear flux scale, whereas the MHD spectra are on a non-linear flux scale. This difference could be removed by flux calibrating the photographic plates (e.g. by identifying spectrophotometric standards on the plates). A final problem could be that the synthetic spectra have an unrealistically high S/N. This could lead the networks to lock onto certain weak features in the synthetic spectra that cannot be used as a source of class discrimination in the lower S/N MHD spectra. One way around this would be to add random noise to the synthetic spectra and use each synthetic spectrum several times in the training set with different random noise added in each case. Nonetheless, the narrow distribution of residuals about the median calibrations demonstrate that these potential problems have only a small effect on our technique and on the resulting calibration shown in Figure 5 and Table 2.

Table 2. $T_{\rm eff}$–MK spectral type calibrations from a committee of neural networks trained on [M/H] = −0.2 spectra. These values are the tabulations of those shown in Figure 5; $\tilde{T}_{\rm eff}$ is the median calibration temperature and σ_{68} is its error bar shown in Figure 5. N is the number of stars of each spectral type, and hence the number of points used to define the distribution from which $\tilde{T}_{\rm eff}$ and σ_{68} are determined. Note that σ_{68} measures the spread of this distribution and is *not* the standard error on the median, which is $\sigma_{68}\sqrt{\pi/2N}$.

		Dwarfs (V)				Subgiants (IV)				Giants (III)			
Spectral Type		$\tilde{T}_{\rm eff}$ / K		σ_{68}	N	$\tilde{T}_{\rm eff}$ / K		σ_{68}	N	$\tilde{T}_{\rm eff}$ / K		σ_{68}	N
13	B5	12145	±	685	3								
14	B6												
15	B7	10842	±	301	4					10868	±	173	3
16	B8	11061	±	884	7	11106	±	1276	5	9568	±	1002	3
17	B9	10877	±	1118	21	10695	±	723	13	9806	±	836	6
18	A0	9762	±	651	51	9139	±	779	6	9211	±	538	4
19	A1	9298	±	678	26	9109	±	206	7	8907	±	84	3
20	A2	8862	±	454	7	8190	±	662	8	7919	±	580	7
21	A3	8364	±	309	16	8110	±	182	4	7903	±	285	4
22	A4					8006	±	213	4				
23	A5	7712	±	347	9	7982	±	360	4	7573	±	323	7
24	A6	7711	±	266	6								
25	A7	7591	±	242	8	7503	±	219	3	7628	±	192	4
26	A8	7645	±	259	12								
27	A9	7466	±	311	76	7188	±	217	4				
28	F0	7261	±	295	122	7003	±	187	6				
29	F2	7040	±	281	135	6956	±	132	9	7081	±	207	3
30	F3	6838	±	257	265	6672	±	244	5				
31	F5	6643	±	280	269	6618	±	147	5				
32	F6	6492	±	254	146								
33	F7	6364	±	275	170								
34	F8	6324	±	275	65								
35	G0	6141	±	316	145								
36	G1	6161	±	333	83								
37	G2	6015	±	328	70								
38	G3	5843	±	307	157	6041	±	166	6	5297	±	309	4
39	G5	5613	±	438	116	5598	±	105	3	5193	±	265	10
40	G6	5429	±	373	50	5406	±	389	17	5263	±	329	15
41	G8	5171	±	381	59	4992	±	311	132	5076	±	332	154
42	K0	4914	±	378	31	4847	±	426	20	4903	±	344	387
43	K1	5149	±	586	11	4780	±	268	23	4744	±	282	285
44	K2	4601	±	750	7					4658	±	254	198
45	K3	4662	±	2057	9					4642	±	290	105
46	K4									4736	±	379	53
47	K5									4585	±	328	30

6.2 Metallicity Effects

It is apparent from Figures 3 and 4 that, when the metallicity of the synthetic spectra is decreased, the neural networks give *systematically* lower effective temperatures for stars of a given spectral type. The calibration results from training neural networks on synthetic spectra of [M/H] = 0.0 and of [M/H] = −0.5 confirm this trend. An explanation of this will be discussed shortly. This is an important result because it demonstrates that the network is sensitive to metallicity, and that in principle this sensitivity can be exploited to give metallicity calibrations for the MHD spectra. How this could be done will be discussed later.

The best agreement between our neural network calibrations and the literature calibrations was obtained when the networks were trained on [M/H] = −0.2 spectra. The literature calibrations are of course based on the same original MK standard spectra; hence the metallicity at which an agreement is achieved tells us something about the mean metallicity of the MHD stars and the MK system in general. That this comes out to [M/H] = −0.2 is not surprising, as it is well known that [M/H] = −0.2 for the local disk (see, for example, Wyse & Gilmore, 1995, and references therein to earlier work). If the distances to the MHD stars are calculated and plotted as a histogram, the giants and dwarfs show a peak in the distribution at $\approx 400\,pc$ and $\approx 100\,pc$ respectively (Bailer-Jones 1996). For many of the stars, which are at high Galactic latitudes, these distances are approximately distances out of the Galactic plane. The MHD stars are therefore primarily Galactic thin disk stars. The thin disk has exponential scale heights of $\approx 100\,pc$ and $\approx 300\,pc$ for the young and old stars respectively (Gilmore & Wyse 1985). Fitting a Gaussian to the distribution of stars plotted against [Fe/H], Gilmore and Wyse (1985) reported a mean

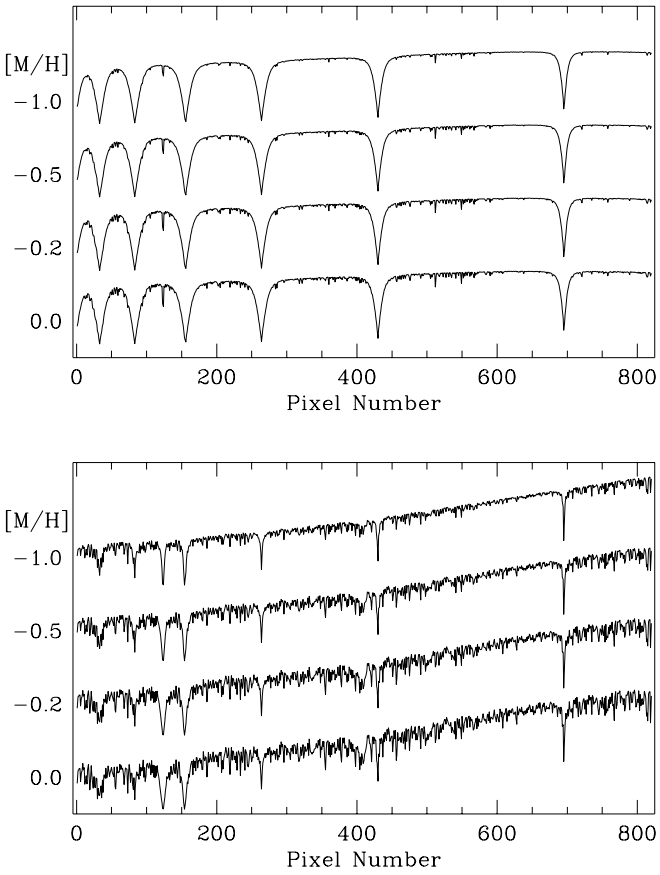

Figure 6. Metallicity effects in the synthetic spectra. The top four spectra are $T_{\rm eff} = 10\,000$ K, $\log g = 4.0$. The bottom four are $T_{\rm eff} = 6250$ K, $\log g = 4.0$. In each case metallicity increases down the box, with the value of [M/H] shown on the left. The spectra have been re-binned to the dispersion of the MHD spectra, but have not been blurred.

abundance $\overline{\rm [Fe/H]} = 0.0$ and $\sigma_{\rm [Fe/H]} = 0.15$ for the young thin disk, and $\overline{\rm [Fe/H]} = -0.3$ and $\sigma_{\rm [Fe/H]} = 0.2$ for the old thin disk. (As SPECTRUM uses standard relative abundances, [M/H] ≈ [Fe/H] for disk stars.) It is therefore plausible that the MHD spectra have an average metallicity of −0.2. It is certainly unlikely to be as high as 0.0 or as low as −0.5 or −1.0. By comparison, the metallicity fit for the Galactic thick disk (scale height $\gtrsim 1\,kpc$) is $\overline{\rm [Fe/H]} = -0.6$ and $\sigma_{\rm [Fe/H]} = 0.3$, and for the Galactic extreme spheroid (scale height $\gtrsim 4\,kpc$) it is $\overline{\rm [Fe/H]} = -1.5$ and $\sigma_{\rm [Fe/H]} = 0.5$.

Why do the $T_{\rm eff}$ calibrations change with metallicity? The strength of the metal lines in a synthetic spectrum increases with increasing metallicity, as can be seen from Figure 6. However, at a fixed metallicity, the apparent number and strength of the metal lines also generally increases as $T_{\rm eff}$ decreases. Suppose a network is trained on metal rich stars and gives the $T_{\rm eff}$ of a certain MHD star as T. A neural network trained on metal weak stars would assign this MHD star a temperature *lower* than T, because this MHD spectrum is closer in appearance to a cooler star in the set of metal weak spectra. In other words, because the neural networks are trained on synthetic spectra of a single metallicity, they are not entirely able to distinguish metallicity features from temperature ones when attempting to classify MHD spectra. We tried to solve this problem by training a network on a set of spectra of all four metallicities (a total of 620 spectra), with the hope that the networks could marginalize over metallicity effects. But this gave even poorer calibrations, both in terms of poorer agreement with the published calibration and a poorer $T_{\rm eff}$–SpT correlation.

This confusion of temperature effects with metallicity ones is not just a problem for neural networks. There is a real correlation between metallicity and temperature. At lower temperatures there is less excitation and ionization of the metals, and hence more metal lines in the spectrum. Although the [M/H] value of the MHD star is constant, it is difficult for the neural network to isolate those features which vary only with metallicity and not with temperature. Thus simply training a neural network on whole spectra with a range of metallicities is unlikely to help, because, rather than permitting the network to 'ignore' metallicity effects, this is more likely to mislead it.

It is clearly essential that metallicity effects are considered when attempting to obtain $T_{\rm eff}$ calibrations for spectra. If the metallicity of the observed spectrum could be determined first, then it could be used as an extra input to a network along with the spectrum. If this network were trained on synthetic spectra with a range of metallicities, then we could imagine that the additional metallicity information would permit an appropriate $T_{\rm eff}$ determination. A more suitable alternative would be to have a series of networks, each trained on spectra of a single metallicity, and then to use the network with the appropriate metallicity to evaluate $T_{\rm eff}$: This is more or less what we did earlier by selecting that metallicity which gave the best agreement with existing $T_{\rm eff}$–MK calibrations. The drawback with these approaches is that they require the metallicity to be known in advance. Jones, Wyse & Gilmore (1995) and Jones et al. (1996) describe a method of determining metallicities from relatively low S/N ratio (10–20) spectra, although at a higher resolution than the MHD spectra. They use spectroscopic indices calibrated by synthetic spectra to measure [Fe/H] and avoid confusion with temperature by using an independent photometric measure of $T_{\rm eff}$.

Despite the confusion between temperature and metallicity, Figures 3 and 4 nonetheless show that metallicity produces a *systematic* rather than a *random* effect on the $T_{\rm eff}$–SpT correlation. In principal, therefore, this metallicity signature can be exploited by the neural networks to obtain metallicities for the MHD stars. The question is how to do it. One solution may to be to determine [M/H] and $T_{\rm eff}$ simultaneously, e.g. by using a neural network with two outputs. A direct approach to determining metallicity independently of temperature may work by identifying only those spectral regions which are most sensitive to metallicity and least sensitive to temperature, such as certain iron lines in late-type stars (Keenan& McNeil 1976). With only these regions as inputs to a neural network, the 'noise' caused by temperature variations is greatly reduced, making it more likely that the neural network could recognise metallicity features. This is an area for future work.

7 SUMMARY

We have shown that neural networks trained on synthetic spectra provide low-error predictions for the effective temperature, $T_{\rm eff}$, of a star based on the star's optical spectrum. By applying these neural networks to spectra with exisiting MK classifications, we have obtained a calibration between the MK spectral type parameter and $T_{\rm eff}$. This calibration shows a good agreement with a number of calibrations from the literature. The calibration was obtained statistically from a number of optical spectra at each spectral type with a resolution of $\approx 2\,\text{Å}\,pix^{-1}$. The precision of this calibration is largely limited by the cosmic scatter in temperature for a given MK class and by limitations of the stellar models. It has also been shown that metallicity effects have to be considered when trying to determine $T_{\rm eff}$. Further work is required before neural networks can be used to accurately quantify metallicities. In particular, further processing of the synthetic spectra into the format of the MHD spectra (e.g. by the addition of noise and calibration of the flux scale) may be required. Our work nonetheless demonstrates that our networks are sensitive to metallicity.

ACKNOWLEDGMENTS

We would like to thank Nancy Houk for kindly loaning us her plate material. We are grateful to Richard Gray for the use of his spectral synthesis program and his effort in converting the code to operate on our computer system. We would also like to thank Robert Kurucz for the use of his model atmospheres.

REFERENCES

Bailer-Jones C.A.L., 1996, PhD thesis, Univ. Cambridge
Bailer-Jones C.A.L., Irwin M., von Hippel, T., 1997a, in preparation
Bailer-Jones C.A.L., Irwin M., von Hippel, T., 1997b, MNRAS, submitted
Bishop C.M., 1995, Neural Networks for Pattern Recognition. Oxford Univ. Press, Oxford
Cawson M.G.M., Kibblewhite E.J., Disney M.J., Phillipps S, 1987, MNRAS, 224, 557
Corbally C.J., Gray R.O., Garrison R.F., eds., 1994, The MK Process at 50 Years. Astronomical Society of the Pacific Conference Series 60. Astronomical Society of the Pacific, San Francisco
Gilmore G., Wyse R.F.G., 1985, ApJ, 90, 2015
Gray D.F., 1992, The Observation and Analysis of Stellar Photospheres. 2nd edn, Cambridge Univ. Press, Cambridge
Gray R.O., Arlt J.S., 1996, Bulletin of the American Astronomical Society, 188, 58.01
Gray R.O., Corbally C.J., 1994, AJ, 107, 742
Hertz J., Krogh A., Palmer R.G., 1991, Introduction to the Theory of Neural Computation. Addison-Wesley, Redwood City
Hertzprung E., 1911, Potsdam Pub., 63, 1
Houk N., Cowley A.P., 1975, University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars. Vol. 1: Declinations -90 to -53 degrees
Houk N., 1978, University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars. Vol. 2: Declinations -53 to -40 degrees
Houk N., 1982, University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars. Vol. 3: Declinations -40 to -26 degrees
Houk N., Smith-Moore M., 1988, University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars. Vol. 4: Declinations -26 to -12 degrees
Houk N., 1994, in Corbally C.J., Gray R.O., Garrison R.F., eds, Astronomical Society of the Pacific Conference Series 60, The MK Process at 50 Years. Astronomical Society of the Pacific, San Francisco, p. 285
Jones J.B., Wyse R.F.G., Gilmore G., 1995, PASP, 107, 632
Jones J.B., Gilmore G., Wyse R.F.G., 1996, MNRAS, 278, 146
Keenan P.C., McNeil R.C., 1976, An Atlas of Spectra of the Cooler Stars: Types G, K, M, S and C. The Ohio State University Press, Ohio
Kurucz R.L., 1979, ApJS, 40, 1
Kurucz R.L. 1992. In Barbuy B., Renzini A., eds, Proc. IAU Symp. 149, Stellar Populations of Galaxies. Kluwer, Dordrecht, p. 225
Lahav O., Naim A., Sodré L., Jr., Storrie-Lombardi M.C., 1996, MNRAS, 283, 207
Miller A.S., 1993, Vistas in Astronomy, 36, 141
Morgan W.W., Abt H.A., Tapscott J.W., 1978, Revised MK Spectral Atlas for Stars Earlier than the Sun. Yerkes Observatory, University of Chicago, Chicago and Kitt Peak National Observatory, Tucson
Morgan W.W, Keenan P.C., Kellman E., 1943, An Atlas of Stellar Spectra with an Outline of Spectral Classification. University of Chicago Press, Chicago
Russell H.N., 1914, Nat, 93, 227
Schmidt-Kaler T.H., 1982, in Schaifers K., Voigt H.H., eds., Landolt-Börnstein, vol. 2b. Springer-Verlag, Berlin.
Storrie-Lombardi M.C., Lahav O., 1994, in Arbib M.A., ed., Handbook of Brain Theory and Neural Networks. MIT Press, Boston
von Hippel T., Storrie-Lombardi L., Storrie-Lombardi M.C., Irwin M., 1994, MNRAS, 269, 97
Wyse R.F.G., Gilmore G., 1995, AJ, 110, 2771