Automated classification and stellar parameterization

Sunetra Giridhar
S. Muneer
A. Goswami

IIA In-House Scientific Meeting
12 April 2007

INDIAN INSTITUTE OF ASTROPHYSICS,
BANGALORE 560034, INDIA

Introduction

The MK spectral types are classical description of stellar spectra. Although the two dimensional MK spectral type (SpT) and luminosity class (LC) are related to temperature and gravity of a star, the SpT are not assigned based upon these parameters but use visual appearance of stellar spectra. The MK classification involves comparing spectra to be classified with those of classification standards of defined class. The advantage of MK system has been that it is model independent and works well even with spectra of modest resolution. N. Houk and her collaborators have done a monumental work of determining SpT and LC for all stars in HD catalog (~ 12,000 stars up to Vmag ~ 11) with RMS error of 0.6 in SpT and 0.25 in LC. This data has been used as reference for automated classifications (see for example von Hippel et al. 1994, Bailer-Jones et al. 1998).

Methods of automated spectral classification

The most commonly used automated spectral classification methods are based on (a) Minimum Distance Method (MDM) (b) Gaussian Probability Method (PDM), (c) Principal Component analysis (PCA) and (d) Artificial Neural Network (ANN). Quantitative methods involving measurement of equivalent widths of certain lines, line strength ratios etc and calibration of these quantities in terms of stellar parameters have also been used. For example, Stock and Stock (1999) used equivalent widths of 19 absorption lines (B-V) colors and Mv derived from Hipparcos catalog for a sample of 487 stars for calibration of M_V . Their algorithms can predict M_V from these line strengths for all spectral types with an average error of 0.26 mag.

MDM

The classification is done by minimizing distance metric between the object to be classified and each member of a set of templates. The object is assigned the class of the template, which gives the smallest distance. If the star spectrum is represented by a vector X = $(x_1, x_2 \ldots\ldots x_i, \ldots\ldots\ldots x_N)$ and template c is represented by another vector S = $(s_1, s_2 \ldots\ldots s_i, \ldots\ldots\ldots s_N)$

the distance D*c* is evaluated $$D_c = \frac{1}{N}\left[\sum_{i=1}^{i=N} w_i^{\{c\}} \left| x_i - s_i^{\{c\}} \right|^p\right]^{1/p}$$

where $w_i^{\{c\}}$ is the weight assigned to spectral element i of the class c.The spectrum X is assigned class c for which D_c is minimum. The weights are assigned to spectral elements based on their relative importance in determining the spectral class. In this approach the number of templates used to define subclasses limit the accuracy of classification. Interpolation can be made to make intra-class assignment.

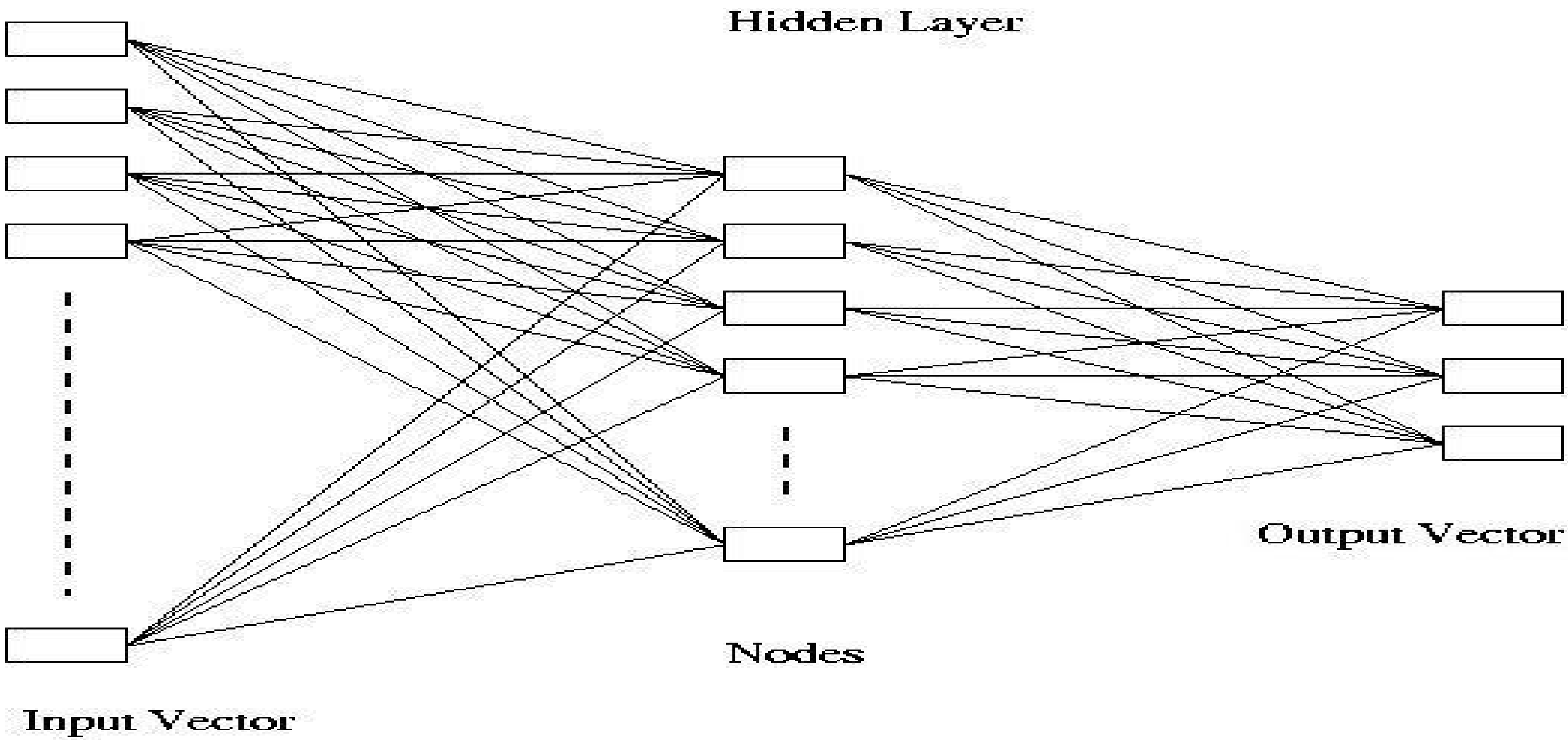
Katz et al. (1998) used this method with χ^2 weighing on high resolution Elodie spectra using a large number of reference stars of known T_{eff} , log *g* and [M/H] to derive atmospheric parameters of target stars. These authors achieved accuracy of 86 K in T_{eff}, 0.28 in log g and 0.35 in [M/H]. Vansevicius and Bridzius (1994) used MDS with χ^2 weighing to estimate SpT and M_V from Vilnius photometric indices. An accuracy of 0.7 was achieved for SpT and 0.8mag for M*V* over spectral type range O5 to M5.

Principal Component Analysis

It is a method of representing a set of N dimensional data by means of their projection onto a set optimally defined axes. Since these axes (Principal components) form an orthogonal set, a linear transformation of the data is achieved. Not all components are important. Components that represent large variance are important while those represent least variance can be ignored and data set can be replaced by significant components alone resulting in reduction of the data size. These compressed data sets are used as input for neural networks. Bailer-Jones et al. (1998) had demonstrated that precise calibration could be done using these compressed spectra and that the optimal compression also results in noise removal. Singh et al. (2006) have used a variation of PCA technique to restore missing data in a sample of 300 stars in Indo-US coude feed spectral library.

Neural Network

As explained very well in numerous papers of Bailer-Jones, Ted von Hippel and others, it is a computational method which can provide non-linear parameterized mapping between an input vector (a spectrum for example) and one or more outputs like SpT, LC or T_{eff}, log *g* and [M/H]. The model is generally supervised; it means that for the network to give required input-output mapping, it must be trained with the help of representative data patterns. These are stellar spectra for which classification or stellar parameters are well determined. The training procedure is a numerical least square error minimization method. The training proceeds by optimizing the network parameters (weights) to give minimum classification error. Once the network is trained the weights are fixed, the network can be used to produce output SpT, LC or T_{eff}, log *g* and [M/H] for an unclassified spectrum.



Neural Network Configuration

Neural Network

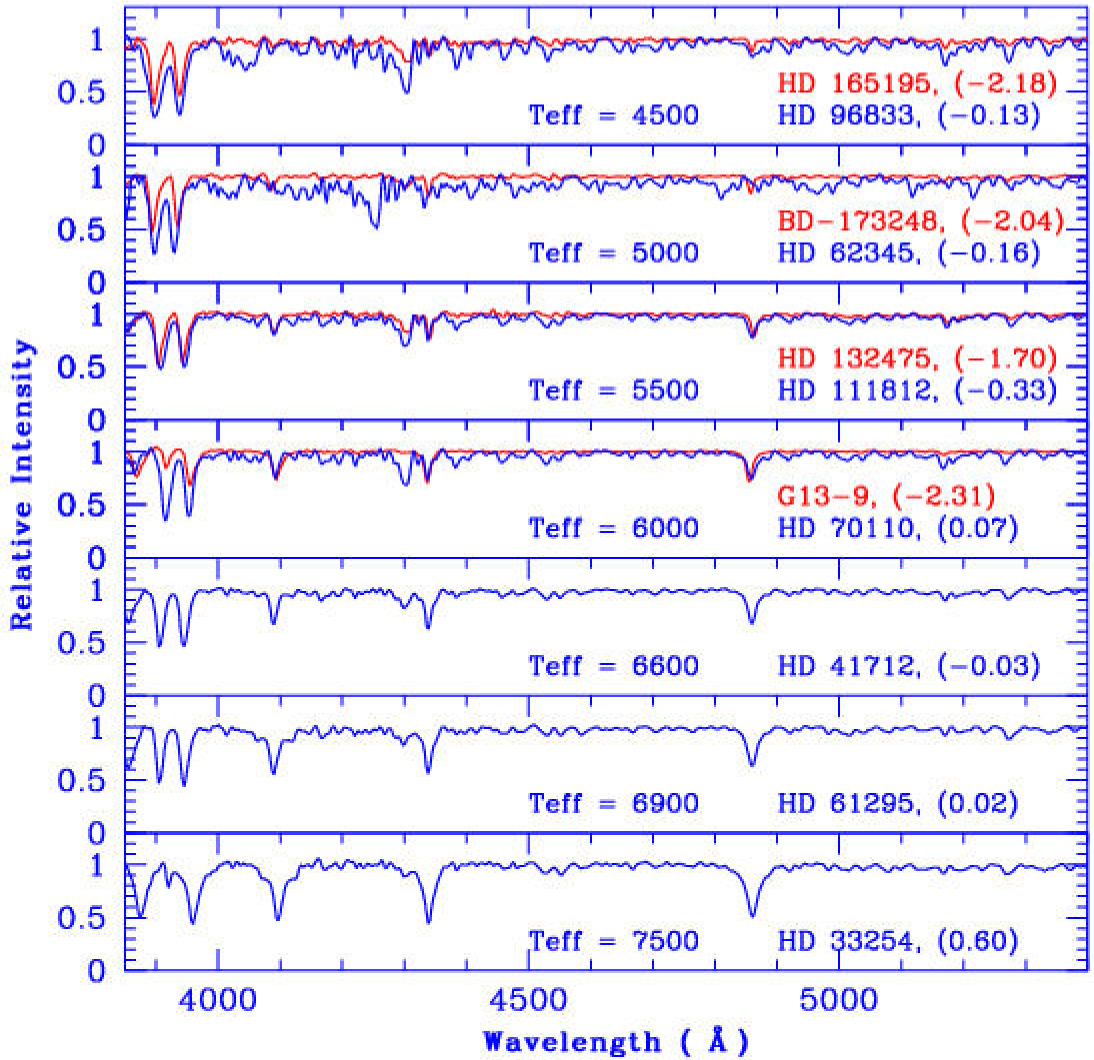
As shown in figure 1, the neural network has one input layer containing stellar spectrum. Each of the input nodes connects to every node in the next layer of nodes called the hidden layer. The neural network architecture may contain one or more hidden layers. Each of these connections has a weight w_{ij} associated with it. The jth node in hidden layer forms a weighted sum of its inputs. It then passes this sum through a non-linear signoid transfer function to give final output from this node. The outputs from nodes in the hidden layer serve as input to the node in the output layer, which again forms weighted sum of its inputs. The training takes place as follows. The weights are initially set with random values over a small range. When the spectrum is fed into the network, the output would also be random. By comparing this output with the target output we can adjust the weights to give an output that is closer to the target value. 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 value. The network training is performed by minimizing the least square error. Since the output from neural network is some non-linear function of all of the network input, it implies that the network output is based upon the appearance of whole spectrum. Depending upon the training data the network will learn which wavelength features are more significant than others in determining the correct spectral parameters and correspondingly would assign appropriate values to the network weights. Here the weights are updated backwards from the output layer through the hidden layer hence the algorithm is called back-propagation method.

Applications of Neural Network

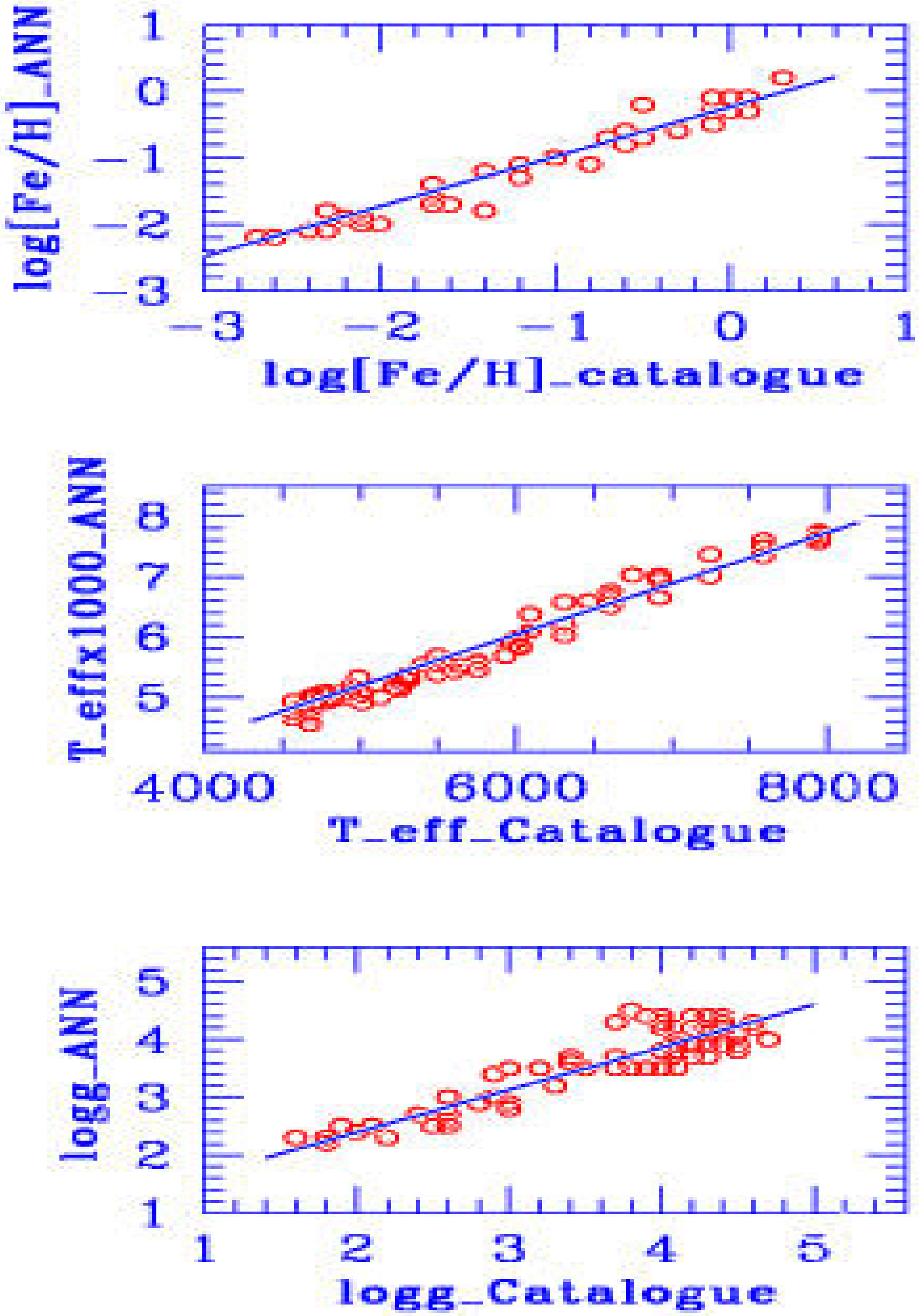
The ANN has been used in very large number of stellar applications. Vieira and Ponz (1995) have used ANN on low-resolution IUE spectra and have determined SpT with an accuracy of 1.1 subclass. Although they attempted classification also with MDM, the errors of classification were larger than that of ANN. In visual region Bailer-Jones, Irwin and von Hippel (1998) used ANN to classify spectra from Michigan Spectral Survey with an accuracy of 1.09 SpT. Visual-near IR spectra were classified by Weaver & Torres-Dodgen(1997) using a two step approach. At first a coarse classification is being done to get main spectral class say F, then it is further classified by more specialist network for that class. This approach results in an accuracy of 0.4 to 0.8 for SpT and 0.2 to 0.4 in LC. Allende Prieto et al. (2000) used it in their search of metal-poor stars. Snijder et al. 2001 used ANN for the three dimensional classification of metal-poor stars. Willemsen et al. (2005) have used ANN to estimate metallicity for main sequence turn-off, subgiants and red giant stars in the globular clusters M55 and / Centauri using the medium resolution spectra of cluster members.

Parameterization of OMR spectra

We have made a modest effort to use ANN to parameterize a sample of stars in temperature range 4500 to 8000 K. We have used a medium resolution Cassegrain spectrograph with 2.3m Vainu Bappu Telescope at VBO, Kavalur, India. The spectrograph gives a resolution R (~ 1000) when used with a grating of 600 grooves/mm and a camera of focal length 150mm. The detector used is 1K X 1K CCD based on Thomson TH77883 chip. The spectral coverage is 3800-6000 Å. The pre-processing of spectra was carried out following a procedure very similar to those of Snijder et al. 2001. We have observed stars from the list of Allende Prieto and Lambert (1999), Alonso, Arribas Martinez-Roger (1996), Snijder et al. (2001) to develop a library of stars with known temperatures, gravities and metallicities. These spectra were used for training and testing the network. We have also observed stars from the lists of metal-poor candidates to estimate the metallicity for them. Although more than 200 stars are observed, the results are reported for 90 stars for which atmospheric parameters T_{eff} and log *g* are well determined, [Fe/H] was known for 47 of them. We have used a software developed by B.D. Ripley based on back propagation technique. Figure shows a few representative spectra. The preliminary results based on 680:11:3 architecture are presented in figure 3. The RMS error for T_{eff} = 200 K, [Fe/H] = 0.3 dex, and that of log g = 0.4 dex. We propose to experiment with an architecture containing two hidden layers instead of one to further reduce the errors of these estimated parameters.



Relative Intensity
Wavelength (Å)
Teff = 4500
HD 165195, (−2.18)
HD 96833, (−0.13)
Teff = 5000
BD−173248, (−2.04)
HD 62345, (−0.16)
Teff = 5500
HD 132475, (−1.70)
HD 111812, (−0.33)
Teff = 6000
G13−9, (−2.31)
HD 70110, (0.07)
Teff = 6600
HD 41712, (−0.03)
Teff = 6900
HD 61295, (0.02)
Teff = 7500
HD 33254, (0.60)



log[Fe/H]_ANN
log[Fe/H]_catalogue
T_effx1000_ANN
T_eff_Catalogue
logg_ANN
logg_Catalogue

Future Goals

It is very important to envisage an approach that would give quick, reliable spectral classifications (or stellar parameters) for the large body of data for stars falling in all regions of HR diagrams. A single ANN architecture may not give the same desired accuracy over full range of spectral types and luminosity classes. A pilot program using the photometric inputs e.g. Strömgren indices, special photometric indices measuring the strengths of molecular bands for late type stars could serve as preprocessor and help in identifying a set of specialist networks which would lead to classification of desired accuracy. A specialist system needs to be evolved for A-type star for the quick identification of chemically peculiar, magnetic or emission line stars. An expert system should also give strength of α elements or that of carbon using CH, CN bands.

Special network need to be developed for objects displaying complex spectra such as Symbiotic stars, Novae and Supernovae. Here the network must be trained on flux calibrated spectra and must use emission line strength as well as shape and structure of the continuum (composite for symbiotic stars and novae) for classification purposes.

Acknowledgements

It is a pleasure to thank Ted Van Hippel for his help in using the software of B.D. Ripley and general encouragement in different stages of this project.

Table 1 (cont.): Stars observed with OMR Spectrograph

Sl No	Stars	HIP	Vmag	Par	(B-V)	log g	Teff	[Fe/H]	Ref	Obs Dt
47	SAO 61681	47997	10.0	12.85	0.652	4.55	5754.4			07-03-06
48	CD-033337	33221	9.03	9.11	0.48	4.11	5930	-1.40	Snider	04-04-06
49	HD 76932	44075	5.86	46.9	0.51	4.4	5964.8	-0.82	G	05-04-06
50	HD 22484	16852	4.28	72.89	0.57	4.15	5981	-0.11	G,cat	27-02-06
51	G11-44	59376	11.08	4.76	0.43	4.18	6010	-2.07	Snider	04-04-06
52	HD 90860	51414	7.01	10.75	0.622	3.74	6025.6			08-03-06
53	HD 70110	40858	6.18	24.50	0.607	4.01	6025.6	0.07	G,cat	28-02-06
54	HD 76617	44103	8.17	11.02	0.596	4.12	6025.6			05-03-06
55	HD 129401	72041	8.68	10.79	0.607	4.26	6025.6			06-05-06
56	HD 101165	56795	9.18	10.22	0.615	4.34	6025.6			08-03-06
57	HD 23650	17887	9.01	16.13	0.582	4.55	6025.6			07-03-06
58	G13-9	59109	9.99	5.75	0.41	4.24	6082	-2.31	Snider	04-04-06
59	HD 84937	48152	8.28	12.44	0.41	4.00	6090	-2.34	Z,cat	12-03-04
60	HD 107700	60351	4.78	11.93	0.515	3.95	6203.6	-0.06	Meas	08-03-06
61	HD 83808	47508	3.52	24.12	0.516	3.23	6309.6			06-03-06
62	HD 130169	72455	7.13	12.79	0.521	3.93	6309.6			06-05-06
63	HD 65871	39616	8.16	16.92	0.529	4.40	6309.6			06-03-06
64	HD 94028	53070	8.21	19.23	0.498	4.50	6309.6	-1.66	H,cat	08-03-06
65	HD 41712	29002	6.94	11.61	0.455	4.25	6463.0	-0.03	Meas	05-03-06
66	HD 126354	70576	4.33	10.38	0.434	3.01	6606.9			06-05-06
67	HD 91948	52064	6.77	14.17	0.465	3.99	6606.9	-0.03	Meas	08-03-06
68	HD 37613	26996	7.84	10.66	0.455	4.20	6606.9			06-03-06
69	HD 89086	50364	7.62	13.33	0.468	4.22	6606.9			06-03-06
70	HD 21925	16479	8.30	13.59	0.418	4.42	6606.9			07-03-06
71	HD 61295	37339	6.16	11.86	0.374	3.70	6918.3	0.02	E,cat	05-03-06
72	HD 127739	71115	5.91	19.44	0.391	4.02	6918.3	0.08	I,	05-04-06
73	HD 97336	54741	8.15	11.84	0.357	4.35	6918.3			08-03-06
74	SAO 58437	27361	9.19	10.49	0.372	4.40	6918.3			06-03-06
75	HD 58431	36059	7.84	10.73	0.331	4.31	7244.4	-0.07	Meas	05-03-06
76	HD 62196	37802	7.67	15.29	0.313	4.37	7244.4			05-03-06
77	HD 33254	23983	5.43	18.54	0.249	4.04	7585.8	0.60	K	06-03-06
78	HD 27045	19990	4.93	34.87	0.259	4.30	7585.8			06-03-06
79	HD 85844	48590	8.23	10.79	0.263	4.37	7585.8			08-03-06
80	HD 29140	21402	4.25	21.68	0.184	3.81	7943.3	0.27	L	05-03-06
81	HD 56221	35341	5.87	12.18	0.181	3.94	7943.3			06-03-06
82	HD 23190	17575	6.83	11.43	0.210	4.20	7943.3			06-03-06
83	HD 76582	44001	5.68	20.30	0.209	4.25	7943.3			06-03-06
84	HD 43771	30275	7.43	10.93	0.209	4.33	7943.3			05-03-06
85	HD 43750	30165	7.44	11.72	0.201	4.34	7943.3			05-03-06
86	HD 34500	24730	7.41	11.97	0.204	4.36	7943.3			06-03-06

Table 1: Stars observed with OMR Spectrograph

Sl No	Stars	HIP	Vmag	Par	(B-V)	log g	Teff	[Fe/H]	Ref	Obs Dt
1	HD 165195	88527	7.34	2.20	1.29	1.45	4507	-2.18	HH,cat	05-04-06
2	HD 33419	24041	6.11	10.37	1.098	2.50	4570.9			06-03-06
3	HD 37984	26885	4.90	10.80	1.144	2.21	4570.9	-0.55	A, cat	05-03-06
4	SAO 80038	40399	9.77	12.56	1.200	3.45	4570.9			06-03-06
5	HD 96833	54539	3.00	22.21	1.144	2.08	4570.9	-0.13	A, cat	27-02-06
6	HD 107610	60305	6.33	10.61	1.115	2.61	4570.9			08-03-06
7	BD+092870	69746	9.45	2.24		1.62	4672	-2.39	Snider	04-04-06
8	HD 27174	20334	8.25	12.52	1.071	3.43	4677.4			07-03-06
9	HD 34303	24665	6.85	11.64	1.061	2.85	4677.4			06-03-06
10	HD 89962	50851	6.06	14.22	1.119	2.90	4677.4			06,08-03-06
11	HD 100006	56146	5.54	10.03	1.056	2.41	4677.4	+0.02	D, cat	08-03-06
12	HD 107325	60170	5.52	20.78	1.091	3.04	4677.4			27-02-06
13	HD 122563	68594	6.20	3.76	0.90	1.61	4687	-2.62	Snider	04-04-06
14	HD 107752	60387	10.07	0.46	0.75	2.07	4710	-2.74	Snider	04-04-06
15	HD 063791	38621	7.92	1.75		1.80	4750	-1.65	GG cat	05-04-06
16	BD-185550	98339	9.35	1.90	0.92	1.87	4785	-2.89	Snider	05-04-06
17	HD 587	840	5.85	18.18	0.973	3.05	4786.3	-0.24	A, cat	01-02-06
18	HD 134440	74234	9.44	33.68	0.85	4.70	4790	-1.43	DD cat	05-04-06
19	HD 044007	29992	8.06	5.17	0.84	2.00	4830	-1.71	CC cat	05-04-06
20	HD 102070	57283	4.72	9.31	0.97	2.57	4870	-0.11	A, cat	05-04-06
21	HD 115772	65047	9.63	2.11	0.84	2.56	4930	-0.70	Snider	04-04-06
22	HD 87140	49371	9.00	4.38	0.70	2.6	4940.8	-2.02	H	05-04-06
23	BD+173248	85487	9.37	3.67	0.66	1.80	4990	-2.04	FF cat	05-04-06
24	BD+173248	85487	9.37	3.67	0.66	2.94	4995	-2.03	Snider	05-04-06
25	HD 5516	4463	4.40	13.44	0.940	2.58	5011.9	-0.54	A, cat	01-02-06
26	HD 73764	42528	6.60	12.32	0.899	3.22	5011.9			05-03-06
27	HD 118253	66381	7.58	11.48	0.875	3.47	5011.9			06-05-06
28	HD 104163	58502	8.48	11.14	0.879	3.68	5011.9			08-03-06
29	HD 166161	88977	8.16	3.25	0.98	1.84	5125	-1.22	Snider	05-04-06
30	G13-38	60747	10.51	10.95	0.71	4.60	5220	-0.96	Snider	04-04-06
31	HD 41116	28734	4.16	21.64	0.835	2.97	5248.1	-0.01	A,cat	05-03-06
32	HD 233608	45098	9.40	13.84	0.879	4.34	5248.1			07-03-06
33	HD 148408	80630	9.62		0.71	4.5	5260	-0.8		05-04-06
34	G60-46		11.00			4.59	5300	-1.19	Snider	04-04-06
35	HD 108317	60719	8.04	4.53		3.33	5310	-2.27	Snider	04-04-06
36	G43-5		12.52		0.65	4.66	5310	-2.12	Snider	04-04-06
37	G141-19	90957	10.55	3.57	0.64	4.00	5400	-2.30	JJ,cat	05-04-06
38	HD 69960	41022	8.00	16.34	0.756	4.06	5495.4			28-02-06
39	HD 76909	44137	7.84	20.98	0.756	4.22	5495.4			28-02-06
40	HD 76218	43852	7.69	38.21	0.771	4.59	5495.4			28-02-06
41	HD 132475	73385	8.57	10.85	0.59	3.76	5550	-1.70	Snider	04-04-06
42	HD 149996	81461	8.49	14.37	0.62	4.1	5600	-0.65	EE,cat	05-04-06
43	HD 104800	58843	9.22	15.97	0.59	4.3	5630.	-0.6		04-04-06
44	HD 111812	62763	4.93	10.62	0.681	2.93	5754.4	-0.33	F	27-02-06
45	HD 26749	19767	6.74	27.98	0.677	4.11	5754.4			07-03-06
46	HD 95364	53851	8.62	12.24	0.690	4.20	5754.4			08-03-06