A&A 391, 397–406 (2002)
DOI: 10.1051/0004-6361:20020830
© ESO 2002

Astronomy & Astrophysics

Statistical methods of automatic spectral classification and their application to the Hamburg/ESO Survey

N. Christlieb[1], L. Wisotzki[2], and G. Graßhoff[3]

[1] Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany
e-mail: nchristlieb@hs.uni-hamburg.de
[2] Institut für Physik, Universität Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany
e-mail: lutz@astro.physik.uni-potsdam.de
[3] Institut für Philosophie der Universität Bern, Länggassstrasse 49a, 3012 Bern, Switzerland
e-mail: gerd.grasshoff@philo.unibe.ch

Received 10 April 2002 / Accepted 4 June 2002

Abstract. We employ classical statistical methods of multivariate classification for the exploitation of the stellar content of the Hamburg/ESO objective prism survey (HES). In a simulation study we investigate the precision of a three-dimensional classification ($T_{\rm eff}$, $\log g$, [Fe/H]) achievable in the HES for stars in the effective temperature range 5200 K $< T_{\rm eff} <$ 6800 K, using Bayes classification. The accuracy in temperature determination is better than 400 K for HES spectra with $S/N > 10$ (typically corresponding to $B_J < 16.5$). The accuracies in $\log g$ and [Fe/H] are better than 0.68 dex in the same S/N range. These precisions allow for a very efficient selection of metal-poor stars in the HES. We present a minimum cost rule for compilation of complete samples of objects of a given class, and a rejection rule for identification of corrupted or peculiar spectra. The algorithms we present are being used for the identification of other interesting objects in the HES data base as well, and they are applicable to other existing and future large data sets, such as those to be compiled by the DIVA and GAIA missions.

Key words. surveys – methods: data analysis – stars: fundamental parameters – Galaxy: halo

1. Introduction

Ever since powerful computers and digital spectra have become available, there have been efforts to develop algorithms for automatic spectral classification (for a review on the early works see Kurtz 1984). The advantages of automated procedures as compared to manual classification are obvious. First of all, only a few experts are able to perform accurate manual classifications, and it was therefore sought to "freeze" this expert knowledge into computer programs. Such programs would allow to obtain *objective* classifications by *quantitative* criteria, and much larger data sets could be processed than by manual classification. The latter issue has become ever more demanding, with upcoming survey missions like DIVA[1], NGST[2], or GAIA[3]. With each of these satellites, it is planned to detect millions of objects, or even one billion objects in the case of GAIA.

In the last decade, much progress was made in the field of automatic spectral classification, and it was demonstrated that computers are actually capable of performing this task (for a recent, comprehensive review see Bailer-Jones 2001). Using Kurtz' metric distance approach (Kurtz 1984), LaSala (1994) automatically classified digitized objective prism spectra from Houk's plates, with good results (σ = 1.14 MK-types). Penprase (1994) used a similar approach, and applied it to slit spectra with similar spectral resolution and a slightly larger wavelength coverage (see Table 1 for a comparison of the data used, and results obtained). The spectral type accuracy he reached for B0–F5 stars was a bit worse than that of LaSala; i.e., σ = 1.5 MK-types. However, as we will see below, it is very difficult to compare the performance of classification algorithms based on the results published in the literature, because (a) rarely ever is the signal-to-noise ratio (S/N) of the data documented, and the achievable classification accuracy depends critically on S/N; (b) different wavelength ranges and spectral resolutions were used; and (c) the algorithms were applied to stars in differing ranges of spectral type.

The influence of the latter on the achievable classification accuracy is nicely demonstrated by comparing the results of Weaver & Torres-Dodgen (1995) with those of Weaver & Torres-Dodgen (1997). In the former paper, the authors report on supervised automatic classification of stars of spectral type

Send offprint requests to: N. Christlieb,
e-mail: nchristlieb@hs.uni-hamburg.de

[1] http://www.ari.uni-heidelberg.de/diva/

[2] http://ngst.gsfc.nasa.gov/

[3] http://astro.estec.esa.nl/GAIA/

Article published by EDP Sciences and available at http://www.aanda.org or http://dx.doi.org/10.1051/0004-6361:20020830

Table 1. Comparison of automatic spectral classification performances.

Method	Type of spectra	λ range	Disp.	S/N	Types	$\sigma_{\rm type}$	$\sigma_{\rm LC}$	Reference
PCA	Slit/photoelectric	3500–4000 Å	10 Å/px		A0–G0	1.16	0.85	W83
Metric dist.	Slit/CCD	3800–5190 Å	67 Å/mm		F8–G8	0.4		LS94
Metric dist.	Slit/CCD	3500–5100 Å	1–2 Å/px		B0–F5	1.5		P94
ANN	IUE	1150–3200 Å	2 Å/px		O3–G5	1.11		VP95
Metric dist.	IUE	1150–3200 Å	2 Å/px		O3–G5	1.38		VP95
ANN	Slit/Reticon	5750–8950 Å	7 Å/px		A0–A9	0.42*	0.15*	WTD95
ANN	Slit/Reticon	5750–8950 Å	7 Å/px		O4–M6	1.26*	0.38*	WTD97
ANN+PCA	Slit/CCD	3510–6800 Å	5 Å/px		O–M	2.34		SGG98
Manual	Objective prism, widened	3800–5190 Å	108 Å/mm	>100?	B2–M7	0.6**	0.25**	H75–88
Metric dist.	Digitized objective prism	3800–5190 Å	1–3 Å/px	>100?	B	1.14		LS94
ANN	Digitized objective prism	3800–5190 Å	1–3 Å/px	>100?	B2–M7	0.82***		BJIvH98
ANN	Slit/CCD	3850–4450 Å	0.65 Å/px	>20	F5–K5	0.57–0.64		Setal01
Bayes	Digitized objective prism	3200–5300 Å	7–18 Å/px	10–30	F2-K0	< 1.6	< 0.55	This work

References: W83 = Whitney (1983); LS94 = LaSala (1994); P94 = Penprase (1994); VP95 = Vieira & Ponz (1995); WTD95 = Weaver & Torres-Dodgen (1995); WTD97 = Weaver & Torres-Dodgen (1997) SGG98 = Singh et al. (1998); Houk75–88 = Houk (1975), Houk (1978), Houk (1982), Houk & Smith-Moore (1988); BJIvH98 = Bailer-Jones et al. (1998); Setal01 = Snider et al. (2001).

* Mean absolute deviation.

** According to von Hippel et al. (1994).

*** 68% quantile.

A0–A9 with a multi-layer artificial neural network (ANN) with one hidden layer, trained with a back-propagation algorithm. They reached a mean absolute deviation of 0.42 spectral types and 0.15 luminosity classes. In the second paper, the ANN was applied to stars in the range O4–M6, and the mean absolute deviations were only 1.26 spectral types and 0.38 luminosity classes. The results of Weaver & Torres-Dodgen have also shown that spectral classification in the near infrared can be done with the same accuracy as in the "classical" MK spectral range, with spectra of much lower resolution. The resolution used by Weaver & Torres-Dodgen was only 7 Å per pixel, and their spectral range 5750–8950 Å. Their results are comparable to that achieved by others at three times higher spectral resolution in the optical or UV.

To continue with our brief review, in recent years, ANNs have been successfully used for supervised automatic spectral classification by a couple of groups. All of them used multilayer back-propagation networks (MBPNs). Vieira & Ponz (1995) automatically classified spectra of O3–G5 stars obtained with the International Ultraviolet Explorer (IUE; dispersion 2 Å per pixel) with an MBPN. The $1\,\sigma$ error was 1.11 spectral types. They found their ANN classification to be superior to a classification with a metric distance method (σ = 1.38 types). The data used by Singh et al. (1998) were optical (3500–6800 Å) slit spectra with a dispersion of 5 Å per pixel. They used Principal Component Analysis (PCA) to preprocess their spectra, and reduce the number of input nodes. They obtained an accuracy of 2.34 spectral types over the full MK range (O–M). Bailer-Jones et al. (1998) used again Houk's plate material, digitized with the APM plate scanner, yielding a wavelength range of 3800–5190 Å and a dispersion of 1–3 Å per pixel. Their best ANN configuration classified these spectra with an error distribution having a 68% quantile of 0.82 types, and the luminosity classification was correct for 95% of the test sample spectra.

Recently, Snider et al. (2001) used a MBPN for derivation of the stellar parameters $T_{\rm eff}$, $\log g$ and [Fe/H] from moderate resolution (0.65 Å per pixel) spectra. Although their aim is to assign *continuous* parameter values to each spectrum, while we as well as the above mentioned authors carried out *discrete* classifications, we include their work in our review because Snider et al. applied their technique to metal-poor stars, which is also the object type we are mainly concerned with in this paper. Snider et al. report classification accuracies of $\sigma_{T_{\rm eff}} = 135$–150 K, $\sigma_{\log g} = 0.25$–0.30 dex and $\sigma_{\rm [Fe/H]} = 0.15$–0.20 dex. However, it appears from the upper panel of their Fig. 4 that subgiants and horizontal branch stars have been excluded from the sample of stars they studied. A rough graphical analysis of their Fig. 4 reveals that unlike in real samples of stars emerging e.g. from wide-angle spectroscopic surveys, which *do* contain subgiants and horizontal-branch stars, their sample can be classified in $\log g$ with a similar precision by dividing it into two classes "by hand", that is, assigning $\log g = 2.5$ to all stars with $T_{\rm eff} < 5000$ K, and $\log g = 4.5$ to all stars with $T_{\rm eff} > 5000$ K. Furthermore, it is questionable that there is any feature present in their set of spectra which does allow for a gravity classification, since they used continuum divided spectra. The Balmer jump, which is a gravity indicator in cool stars, is therefore removed. In conclusion, while Snider et al. succeeded in using ANNs for automated classification in $T_{\rm eff}$ and [Fe/H], it remains to be demonstrated with a realistic sample that rectified moderate-resolution spectra indeed contain the information needed for a useful gravity classification.

ANN techniques and "classical" statistical methods such as Bayes and minimum cost rule classifications often perform equally well, in terms of e.g. minimizing the total number of misclassifications. In the present work, we employ statistical methods, because their mathematical properties are well-studied, and the formulation of classification rules in the framework of mathematical statistics makes them very transparent.

Before we go into details of the methods we developed (Sect. 3), we give a brief overview of the Hamburg/ESO Survey (HES) in Sect. 2, for better readibility. In Sect. 4 we investigate the classification performance for stars in the effective temperature range 5200 K $< T_{\rm eff} <$ 6800 K achievable in the HES, by a simulation study. We summarize our conclusions in Sect. 5.

2. The Hamburg/ESO Survey

The HES (Wisotzki et al. 1996, 2000) is an objective-prism survey designed to select bright ($12.5 \gtrsim B_J \gtrsim 17.5$) quasars in the southern extragalactic sky ($\delta < +2.5°$; $|b| \gtrsim 30°$). It is based on IIIa-*J* plates taken with the 1 m ESO Schmidt telescope and its 4° prism. The plates were digitized at Hamburger Sternwarte. The HES spectra cover a wavelength range of 3200 Å $< \lambda <$ 5200 Å and have a seeing-limited spectral resolution of typically 15 Å at Hγ and 10 Å at Ca II K 3934 Å. This resolution makes it possible to also exploit the stellar content of the survey very efficiently. For HES example spectra see Fig. 1.

The goals of automatic spectral classification in the HES are (a) three-dimensional classification ($T_{\rm eff}$, $\log g$, [Fe/H]) of the total HES data base currently used for the exploitation of the stellar content, consisting of ~4 million spectra, (b) compilation of complete samples of objects of specific classes, and (c) identification of peculiar objects. These goals are similar to those emerging in DIVA and GAIA. Interesting classes of stars that can be found on HES plates include extremely metal-poor halo field stars (Christlieb & Beers 2000), field horizontal branch stars (Christlieb et al., in preparation), carbon stars (Christlieb et al. 2001a), and white dwarfs (Christlieb et al. 2001b). A large data base with spectra of known type is also very useful for cross-identification with surveys in other wavelength ranges, such as FAUST (as has been demonstrated by Brosch et al. 2000), the Two Micron All Sky Survey[4] (2MASS), or the Deep Near Infrared Survey[5] (DENIS). Furthermore, the data from these surveys can be used to extend the feature vectors associated with HES spectra (see below), and improve the automatic classification in the HES.

3. Automatic spectral classification

In order to achieve our classification aims, we need to construct a *decision rule* which allows us to assign a spectrum with feature vector $\boldsymbol{x}$ to one of the $n_{\rm c}$ classes Ω_j, $j = 1 \dots n_{\rm c}$, defined in the specific classification context. That is, we want to carry out a supervised classification, as opposed to unsupervised classification, where the aim is to group objects into classes not defined before the classification process. Methods of unsupervised classification and their application to HES spectra are presented in Hennig & Christlieb (2002).

3.1. Feature space

The HES data base of digital spectra can be represented by feature vectors $\boldsymbol{x}$, consisting of a set of continuous variables x_i, i.e.

$$\boldsymbol{x} = (x_1, \dots, x_d), \tag{1}$$

where d is the number of features used. It is critical for automatic classification to have a set of reliable features at hand. A wide range of spectral features is automatically measured in the digitized HES objective-prism spectra during the data reduction process (cf. Table 2): stellar absorption and emission lines, absorption bands, continuum shape including spectral breaks, bisecting points of spectral density distribution. These features are measured in unfiltered HES spectra with the methods described in Christlieb et al. (2001b).

3.2. Choosing a feature combination

It is necessary to select a subset of the available features for each classification problem, and each S/N level, because of several reasons.

(1) Blended lines, e.g. Hε+Ca H, can confuse the classification.
(2) It is advantageous to exclude redundant features from the set of features used for classification, since the usage of fewer features results in more stable estimates of the parameters of the multivariate normal distributions (see Eq. (3) below).
(3) The optimal feature set can vary with S/N. For instance, at low S/N it can be useful to only use continuum shape parameters and colors for classification, because no stellar lines can be detected reliably anymore.

The evaluation of the suitability of all $2^d - 1$ possible combinations of d available features for a given classification problem is a complex task at first glance. However, it is usually possible to select a set of appropriate features by physical considerations alone. E.g., when it is desired to select metal-poor stars, only those features that are possibly useful as indicators for $T_{\rm eff}$, $\log g$, [Fe/H], and [C/Fe] need to be considered. By means of accuracy considerations and parameter studies, further features can be rejected. Finally, it is also possible to reduce the dimensionality of the feature space by a priori combining redundant features, e.g. the equivalent widths of the Balmer lines to a sum of equivalent widths. The remaining set of features is then evaluated with the methods described in Sect. 4.1.

3.3. Learning sample

For supervised classification, a *learning sample* is needed. For our purposes, we define a learning sample to be a set of n_l objects for which the feature vectors are known,

$$\{\boldsymbol{x}\} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_{n_l}), \tag{2}$$

and for which it is known to what class they belong. These classes can be defined, e.g., by grouping a set of objects according to their stellar parameters (e.g. $T_{\rm eff}$, $\log g$, [Fe/H]), or

[4] http://www.ipac.caltech.edu/2mass/

[5] http://cdsweb.u-strasbg.fr/denis.html

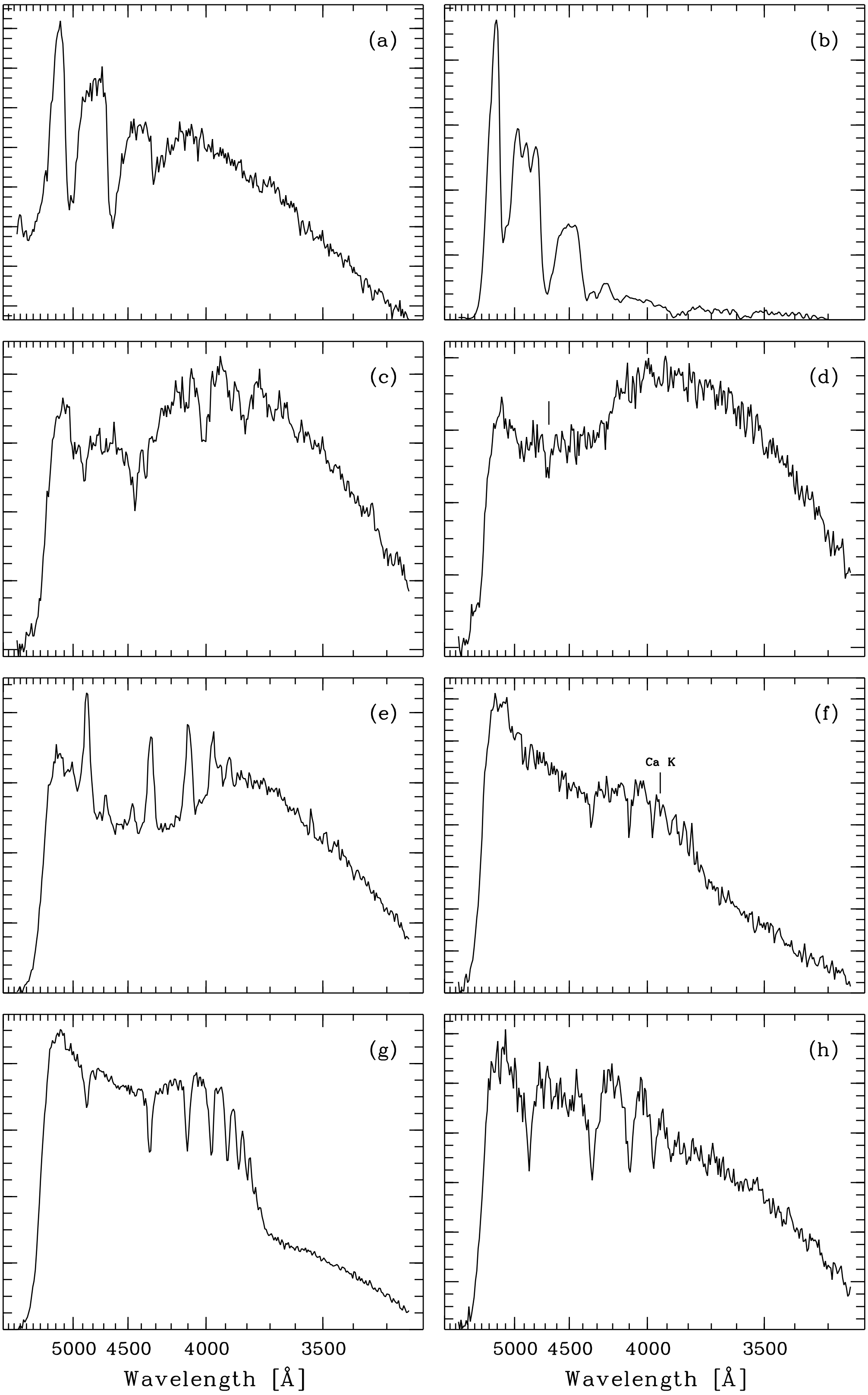

Fig. 1. HES example spectra, illustrating the large variety of object types that can be identified in that survey. Abscissae are wavelength in Å, ordinates are photographic density in arbitrary units. Note that wavelength is *de*creasing from left to right. The sharp drop at ~5400 Å is due to the IIIa-*J* emulsion sensitivity cutoff ("red edge"). Spectra of the following object types are shown: **a)** DQ white dwarf (the red edge of this spectrum is disturbed by an overlapping spectrum); **b)** cool carbon star; **c)** DB white dwarf; **d)** PG 1159 star (the blend of He II $\lambda4686$ and [CIV] $\lambda4660$ is marked); **e)** cataclysmic variable star; **f)** extremely metal-poor star (showing a very weak Ca K line); **g)** FHB/A star; **h)** cool DA white dwarf. The lower two spectra demonstrate that the strength of the Balmer jump can be used as an indicator for the surface gravity $\log g$: While it is invisible in the $\log g \sim 8$ white dwarf, the Balmer jump is pronounced in the $\log g \sim 2$ horizontal branch star.

Table 2. Automatically measured spectral features in the HES. The measurement methods of the features #1–10, #17–20 and #25–27 have been described in Christlieb et al. (2001b) and Christlieb et al. (2001a). The line indices KP, GP and HP were proposed by Beers et al. (1990), and the definition of HP was later refined by Beers et al. (1999). We adapted these indices for the lower resolution of the HES spectra (in particular, the positions of some of the continuum bands were changed), and calibrated them against stars of Beers et al. present on HES plates. The 1 σ scatters of these calibrations are 1.22 Å, 1.41 Å and 1.61 Å, respectively. balmsum can be used to predict HP with an accuracy of σ = 1.55 Å, and this feature is therefore superior to the directly derived HP. From the half power point distances dx_hpp1 and dx_hpp2, $U-B$ and $B-V$ colors can be derived with accuracies of better than 0.1 mag; c_1 can be measured in HES spectra with a precision of 0.15 mag (Christlieb et al. 2001b).

Number	Name	Description	Measurement method
#1	all5160eqw	W_λ of Mg I b triplett/TiO λ 5168	Iterative fit procedure
#2	all4861eqw	W_λ of Hβ	Iterative fit procedure
#3	all4388eqw	W_λ of Fe I λ 4383+85	Iterative fit procedure
#4	all4340eqw	W_λ of Hγ	Iterative fit procedure
#5	all4300eqw	W_λ of G-Band	Iterative fit procedure
#6	all4261eqw	W_λ of Cr I λ 4254 + 75 + Fe I 4260 + 72	Iterative fit procedure
#7	all4227eqw	W_λ of Ca I λ 4227	Iterative fit procedure
#8	all4102eqw	W_λ of Hδ	Iterative fit procedure
#9	all3969eqw	W_λ of Ca H + Hϵ	Iterative fit procedure
#10	all3934eqw	W_λ of Ca K	Iterative fit procedure
#11	balmsum	all4861eqw+all4340eqw+all4102eqw	Meta-feature
#12	CaBreak_sn	S/N Calcium-break	Template matching
#13	CaBreak_cont	Contrast of Calcium-break to continuum	Template matching
#14	KP	Strength of Ca K	Ratio of average pixel values
#15	GP	Strength of G band	Ratio of average pixel values
#16	HP	Strength of Hδ	Ratio of average pixel values
#17	C2idx1	Strength of C_2 λ 5165	Ratio of average pixel values
#18	C2idx2	Strength of C_2 λ 4737	Ratio of average pixel values
#19	CNidx2	Strength of CN λ 4216	Ratio of average pixel values
#20	CNidx3	Strength of CN λ 3883	Ratio of average pixel values
#21	klcomp_1	1. continuum shape coefficient	PCA
#22	klcomp_2	2. continuum shape coefficient	PCA
#23	klcomp_3	3. continuum shape coefficient	PCA
#24	klcomp_4	4. continuum shape coefficient	PCA
#25	dx_hpp1	Half power point distance 1	Summing of pixel values
#26	dx_hpp2	Half power point distance 2	Summing of pixel values
#27	c_1	Strömgren medium band color index	Function of summed pixel values

by manually assigning classes to a set of spectra by comparison with reference objects. With the help of a learning sample, information on the class-conditional probability densities $p(\boldsymbol{x}|\Omega_j)$ can be gained. $p(\boldsymbol{x}|\Omega_j)\mathrm{d}\boldsymbol{x}$ is the probability to observe a feature vector in the range $\boldsymbol{x}\ldots\boldsymbol{x}+\mathrm{d}\boldsymbol{x}$ in the class Ω_j. We inspected the one-dimensional class-conditional probability distributions of the classes covered by the learning samples used in this work, and qualitatively found their shapes to agree well with Gaussians. We hence model $p(\boldsymbol{x}|\Omega_j)$ by multivariate normal distributions, i.e.,

$$p(\boldsymbol{x}|\Omega_j) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma_j|}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)' \Sigma_j^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)\right\}, \quad (3)$$

where j denotes class number, $\boldsymbol{\mu}_j$ the mean feature vector of class Ω_j, and Σ_j the covariance matrix of class Ω_j.

3.4. Decision rules

A central issue in automatic classification is the construction of a decision rule which is optimal for the given classification problem. In the HES, we use three decision rules: the Bayes rule, a minimum cost rule, and a rejection rule.

3.4.1. Bayes' rule

Classification with Bayes' rule minimizes the total number of misclassifications, if the true distribution of class-conditional probabilities $p(\boldsymbol{x}|\Omega_i)$ is used (Hand 1981; Anderson 1984). Using Bayes' theorem,

$$P(\Omega_i|\boldsymbol{x}) = \frac{P(\Omega_i)p(\boldsymbol{x}|\Omega_i)}{\sum\limits_{\forall i} P(\Omega_i)p(\boldsymbol{x}|\Omega_i)}, \quad (4)$$

posterior probabilities $P(\Omega_i|\boldsymbol{x})$ can be calculated. A spectrum of unknown class, with given feature vector $\boldsymbol{x}$, can then be classified using Bayes' rule:

Bayes' rule: *Assign a spectrum with feature vector $\boldsymbol{x}$ to the class with the highest posterior probability $P(\Omega_i|\boldsymbol{x})$.*

3.4.2. Minimum cost rule

In most of the classification problems arising in the HES it is desired to gather a sample of objects of a specific class, or a specific set of classes. In these cases, Bayes' rule is not appropriate, because we do not want to minimize the total number of misclassifications, but the misclassifications between the desired class(es) of objects, and the remaining classes. Suppose we have three classes, A-, F-, and G-type stars, and we want to gather a complete sample of A-type stars. Then only misclassifications between A-type stars and F- and G-type stars (and vice versa) are of interest. More specifically, misclassifications of A-type stars to F- and G-type stars (leading to incompleteness) are least desirable when a complete sample shall be gathered, and erroneous classification of F- and G-type stars as A-type stars (resulting in sample contamination) can be accepted at a moderate rate. Misclassifications between F- and G-type stars can be totally ignored, because the target object type is not involved.

Classification aims like this can be realized by using a minimum cost rule. Cost factors r_{hk}, with

$$0 \leq r_{hk} \leq 1; \qquad h = 1, \ldots, n_{\mathrm{c}}; \quad k = 1, \ldots, n_{\mathrm{c}}, \tag{5}$$

allow to assign relative weights to individual types of misclassifications. The cost factor r_{hk} is the relative weight of a misclassification from class Ω_h to class Ω_k.

Suppose we have an object of unknown class, with feature vector $\boldsymbol{x}$. We ask how large the cost is if it belongs to class Ω_h, and would be assigned to class Ω_k, $h \neq k$. The cost $C_{h\to k}(\boldsymbol{x})$ is:

$$\begin{aligned} C_{h\to k}(\boldsymbol{x}) &= r_{hk} P(\Omega_h|\boldsymbol{x}) \\ &= r_{hk} \frac{P(\Omega_h)\, p(\boldsymbol{x}|\Omega_h)}{\sum\limits_{i=1}^{m} P(\Omega_i)\, p(\boldsymbol{x}|\Omega_i)} = r_{hk} \frac{a_h p_h(\boldsymbol{x})}{\sum\limits_{i=1}^{m} a_i p_i(\boldsymbol{x})}. \end{aligned} \tag{6}$$

In the last step we have used the abbreviations $P(\Omega_h) = a_h$ and $p(\boldsymbol{x}|\Omega_h) = p_h(\boldsymbol{x})$. We do not know to which of the possible classes Ω_h, $h = 1, \ldots, n_{\mathrm{c}}$, the object actually belongs. Therefore, we estimate the expected cost $C_k(\boldsymbol{x})$ for assigning an object with feature vector $\boldsymbol{x}$ to the class Ω_k by computing the following sum of costs:

$$\begin{aligned} C_k(\boldsymbol{x}) &= \sum_{\substack{h=1\\h\neq k}}^{m} C_{h\to k}(\boldsymbol{x}) \\ &= \sum_{\substack{h=1\\h\neq k}}^{m} r_{hk} \frac{a_h p_h(\boldsymbol{x})}{\sum_{i=1}^{m} a_i p_i(\boldsymbol{x})}. \end{aligned} \tag{7}$$

Now we can formulate the minimum cost rule, which minimizes the total cost (Hand 1981).

Minimum Cost Rule: *Assign an object with feature vector* $\boldsymbol{x}$ *to the class* Ω_k *with the lowest expected cost* $C_k(\boldsymbol{x})$.

If the cost factors are chosen such that $r_{hk} \equiv \delta_{hk}$, the minimum cost rule classification is identical to classification using Bayes' rule. In this case the cost for assigning the class Ω_k to a spectrum with feature vector $\boldsymbol{x}$ is the probability that the object belongs to one of the other classes $h \neq k$. This follows immediately from Eq. (7). If $r_{hk} \neq \delta_{hk}$, the total number of misclassifications is *not* minimized, so that the quality of a minimum cost rule classification has to be evaluated by other criteria.

For any given classification aim, one can divide the cost factors to be chosen into three sets:

`t2o`: Cost factor for misclassification of an object of the **t**arget class ("`t`") to ("`2`") one of the **o**ther classes ("`o`").
`o2t`: Cost factor for contamination of the target class.
`o2o`: Cost factor for misclassification between other classes.

Since sample completeness and contamination are interdependent, in practice only the relative value `t2o`/`o2t` has to be adjusted. For this purpose, the classification results as a function of `t2o`/`o2t` are evaluated. The expected error rates, estimated e.g. with the "leaving one out" method (see Sect. 4.1 below), tell which level of completeness and sample contamination will be achieved. Christlieb et al. (1998) presented a software tool for a convenient choice of cost factors.

3.4.3. Rejection rule

Non-mathematically speaking, Bayes' rule assigns the class with the highest relative resemblance to each spectrum to be classified. However, it is ignorant of the absolute resemblance: A spectrum with feature vector $\boldsymbol{x}$ may be assigned to a class with very low posterior probability $p(\Omega_i|\boldsymbol{x})$, if $p(\Omega_i|\boldsymbol{x})$ is even lower for all other classes. This means that a class is assigned to all spectra, even to "garbage spectra" which are disturbed, for instance, by plate artifacts. Therefore, it is useful to apply a rejection rule in addition to either the Bayes rule or the minimum cost rule. The rejection rule can also be used "stand alone" for the identification of peculiar objects, e.g., quasars.

Rejection rule: *Reject an object from classification to class* Ω_i, *if* $A(\Omega_i; \boldsymbol{x}) > \beta$.

The parameter β is a threshold to be chosen, and the parameter A is the *atypicality index* suggested by Aitchison et al. (1977),

$$A(\Omega_i, \boldsymbol{x}) = \Gamma\left\{\frac{d}{2}; \frac{1}{2}\,(\boldsymbol{x} - \boldsymbol{\mu}_i)'\, \Sigma_i^{-1}\, (\boldsymbol{x} - \boldsymbol{\mu}_i)\right\}, \tag{8}$$

where $\Gamma(a; x)$ is the incomplete gamma function and d the number of features used for classification. Use of the above rejection criterion is identical to performing a χ^2 test of the null hypothesis H_0 that an object with feature vector $\boldsymbol{x}$ belongs to class Ω_i at significance level $1 - \beta$, against the alternative hypothesis H_1 that it does belong to class Ω_i. We reject the null hypothesis if its significance level is low, i.e., if it is very unlikely that a feature vector $\boldsymbol{x}$ is observed for class Ω_i, given the multivariate normal distributions (3) are the true distributions of the class-conditional probabilities $p(\boldsymbol{x}|\Omega_i)$.

4. Classification performance

In a first application of automatic spectral classification in the HES, we selected candidates for extremely metal-poor halo field stars. Christlieb (2000) and Christlieb & Beers (2000) have shown that with this method, a very efficient selection of metal-poor stars is feasible. 80% of an investigated sample of 56 highest priority metal-poor candidates were shown by medium-resolution follow-up spectroscopy to have metallicities below −2.0 dex, and results based on a larger sample of

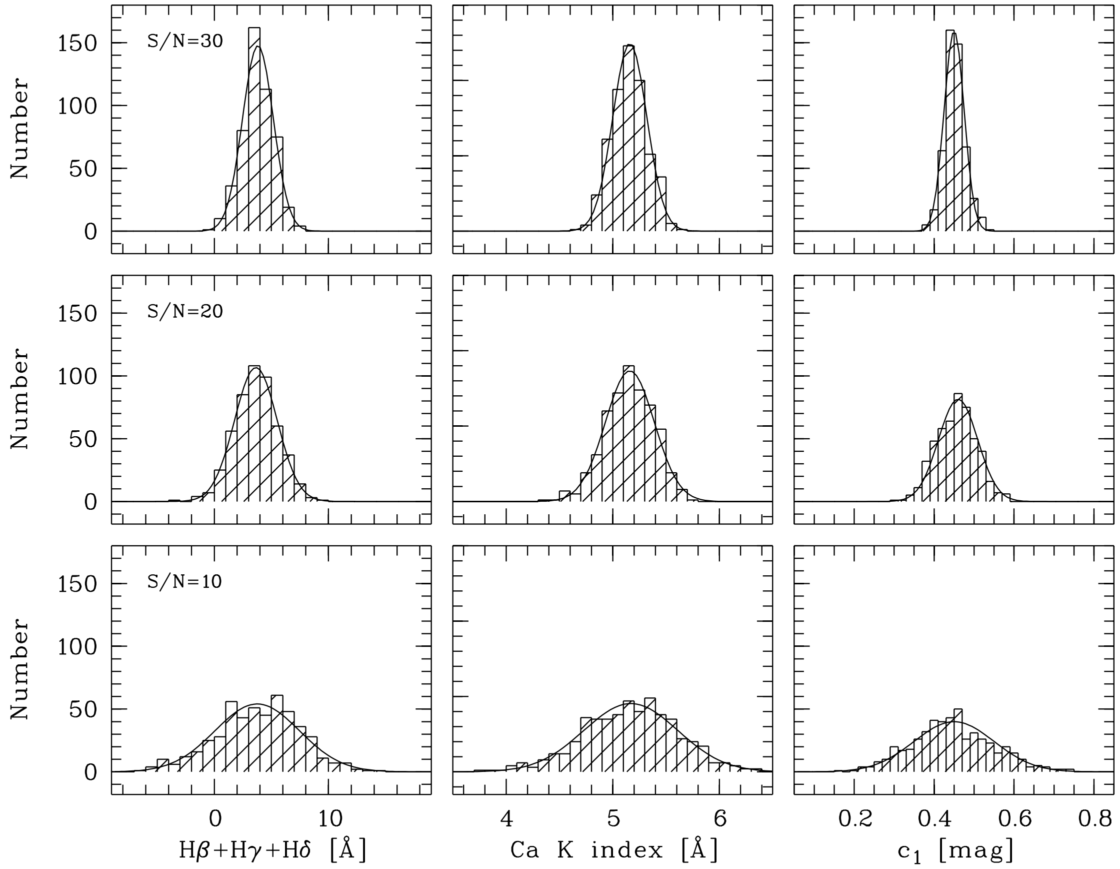

Fig. 2. Distribution of the three features used for classification of main sequence turnoff stars in one learning sample class, and for different signal-to-noise ratios.

stars, including also fainter and lower priority candidates, indicate that the overall efficiency for the selection of stars with [Fe/H] < −2.0 is ∼60% in the HES (Christlieb et al., in preparation). This is the most efficient selection of metal-poor stars ever obtained in a wide-angle survey for such stars. In this paper we focus on results of a systematic investigation of the classification performance for stars in the effective temperature range 5200 K < $T_{\rm eff}$ < 6800 K achievable in the HES, by means of a simulation study.

4.1. Evaluation of classification rules

Classification rules can be evaluated by the number of expected misclassifications (in the case of Bayes' rule), or by the total expected cost (in the case of the minimum cost rule). The three most important methods to estimate these numbers are (Deichsel & Trampisch 1985):

(1) Re-substitution
(2) "Hold out" method
(3) "Leaving one out" method.

Re-substitution means that one uses the learning sample also as test sample. The drawback of this method is that one underestimates the number of expected misclassifications, because a classification rule derived with the help of a finite learning sample is always adapted to the individual composition of the learning sample. Therefore, the estimation of the expected number of misclassifications is biased (Deichsel & Trampisch 1985).

An improvement in this respect is gained when the "hold out" method is used. Here one randomly divides the learning sample disjointly into a new, smaller learning sample, and a test sample. Since the learning sample and test sample are completely independent in this case, an unbiased estimate of the expected error rates is possible (Deichsel & Trampisch 1985). However, the drawback is that one needs a large enough learning sample. When modeling the class-conditional probabilities with multivariate normal distributions, the learning sample size has to be large enough to ensure a robust estimation of the parameters of the distributions.

The problem of learning sample size can be circumvented by using the "leaving one out" method. Suppose we have a learning sample of size $n_{\rm l}$. We exclude object i from the learning sample, and construct the classification rule using the $n_{\rm l} - 1$ remaining objects. Object i is then classified with this classification rule. This procedure is repeated $n_{\rm l}$ times, so that each object of the learning sample is excluded once, and used as test "sample". By adding up the numbers of misclassifications obtained in each step, one gets an unbiased estimate of the expected error rate (Deichsel & Trampisch 1985). The only drawback of this method is that it consumes a lot more computing time than the previously mentioned methods, since n_l classification rules have to be constructed. However, the computing time increases only linearly with learning sample size n_l, so that the usage of the "leaving one out" method was feasible for all

HES learning samples used so far (the largest learning sample used had $n_l = 165\,000$).

4.2. Simulation study on the classification performance in the HES

For our simulation study, we employed a grid of model spectra converted to objective prism spectra with the methods described in Christlieb et al. (2001b). The grid covers the following stellar parameter range:

$T_{\rm eff} = 5200(200)6800\,{\rm K}$

$\log g = 2.2(0.8)4.6$

$[{\rm Fe/H}] = -0.3, -0.9, -1.5(0.3) - 3.6.$

The values in brackets refer to grid point distances. The grid defines 360 classes. Since it is one of the aims of our simulation study to investigate how the classification accuracy changes with S/N, we need to simulate spectra of different S/N. For this, we added Gaussian noise to the grid of simulated spectra so that spectra with $S/N = 5(5)30$ resulted. The adequatness of this noise model for HES spectra has been demonstrated by Christlieb et al. (2001b). It is necessary to produce learning samples for different S/N levels because the width of the class-conditional probability distributions change with S/N (see Fig. 2). We then performed a Bayes classification, using the three features c_1, balmsum (the sum of the equivalent widths of Hβ, Hγ and Hδ), and the Ca K index KP. This feature set was found to be best suitable for the desired three-dimensional classification in parameter studies, and by systematic evaluation of the classification performance of different feature combinations.

For each spectral class 500 simulated spectra were computed, which is a large enough number to randomly subdivide the grid into a learning sample and an independent test sample. To obtain a realistic estimate of the classification performance, two effects have to be taken into account:

Undersampling of the error distribution: In our simulation study, we use a grid of stellar parameters. Therefore, if a classification error is smaller than half of the grid point distance, an error of zero is measured. Therefore, the classification error is systematically underestimated in these cases.

Discretization error: Real samples of stars have a continuous distribution of stellar parameters. These parameters will be mapped to our discrete grid. This results in classification errors for stars having stellar parameters lying between two grid points.

We have taken these effects into account by applying (upward) corrections to the classification errors measured on our model spectra grid (see Fig. 3). For the estimation of the corrections for error distribution undersampling, Gaussian random errors were added to the grid point parameters to simulate classification errors, and the measured errors were compared with the errors known from the chosen σ of the Gaussian distribution. The discretization error correction to be applied was derived by mapping continuously distributed stellar parameters to our grid, and computing the mean difference between real parameters and the parameters detected by the grid.

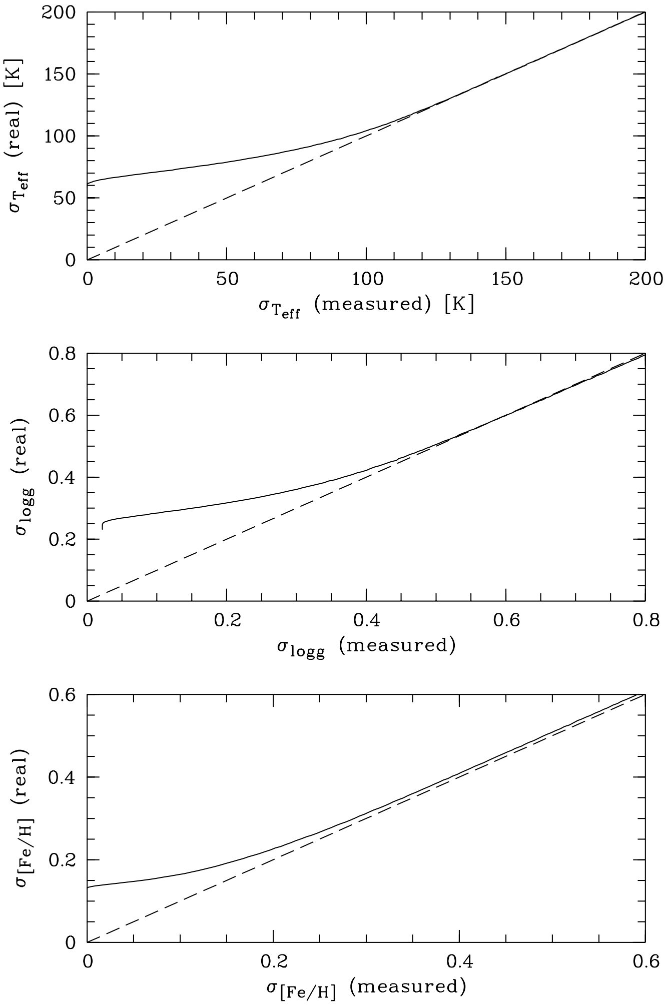

Fig. 3. Corrections applied to the classification errors measured in the simulation study. For explanation see text.

In the stellar parameter range we explored so far, the corrected accuracy in effective temperature classification is better than 400 K for spectra with $S/N > 10$, which typically corresponds to $B_J < 16.5$ (see Christlieb et al. 2001b). The accuracies in $\log g$ and [Fe/H] are better than 0.68 dex for the same magnitude range. Note that the accuracy in [Fe/H] strongly depends on [Fe/H] itself, since the Ca K line, used as metallicity indicator, is not detectable in the spectra of the lowest metallicity turnoff stars at the spectral resolution of the HES. Therefore, a metallicity classification is not possible in that part of the stellar parameter space, resulting in a larger average classification error.

As a plausibility check we compared our results with the classification accuracies we would expect from simple, one-dimensional parametrization approaches using $B - V$, c_1 and KP as temperature, gravity and metallicity indicators, respectively. In the effective temperature range 5200 K $< T_{\rm eff} <$ 6800 K, $\Delta(B - V)/\Delta T_{\rm eff} \sim 0.028$ mag/100 K (Lang 1992). The average accuracy of the HES $B - V$ calibration in the temperature range under consideration, averaged over the full

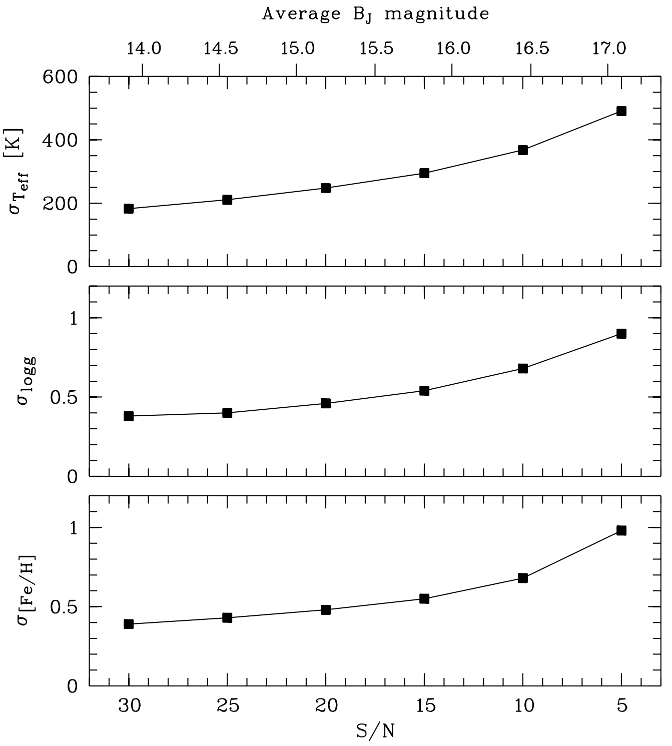

Fig. 4. Classification precision for stars in the effective temperature range 5200 K < $T_{\rm eff}$ < 6800 K in the HES as a function of S/N, as obtained with Bayes classification in our simulation study.

magnitude range covered by the HES, is $\sigma_{B-V} = 0.07$ mag (Christlieb et al. 2001b), so that an average temperature classification accuracy of $\sigma_{T_{\rm eff}} = 260$ K is expected. This is consistent with classification errors of 200–420 K in the magnitude range $14.0 < B_J < 17.5$. $\Delta c_1/\Delta \log g \sim 0.1$–0.3 mag/dex in the effective temperature range under consideration (Lang 1992). The average accuracy of the HES c_1 calibration is $\sigma_{c_1} = 0.15$ mag (Christlieb et al. 2001b), so that we expect a gravity classification precision of $\sigma_{\log g} = 0.5$–1.5 dex. This is consistent with $\sigma_{\log g} < 0.68$ measured in our simulation. Finally, from Fig. 4 of Beers et al. (1990) one can read that at $B - V = 0.5$, the difference in the Ca K index KP between a star of [Fe/H] = −2.0 and a star of [Fe/H] = −3.0 is 2.7 Å and 4.7 Å for dwarfs and giants, respectively. Considering the fact that $\sigma_{\rm KP} = 1.22$ Å in the HES, it is not surprising that classification precisions as high as 0.4 dex can be achieved for the brightest stars in the HES (see Fig. 4).

5. Discussion and conclusions

We have demonstrated that automatic spectral classification of turnoff stars, using "classical" statistical methods, is feasible in the HES with high accuracy. Our results suggest that it might be possible to determine the metallicity distribution function (MDF) of the galactic halo directly from a large sample of HES spectra. The MDF is an important constraint for models of Galactic chemical evolution (see, e.g., Ikuta & Arimoto 1999; Oey 2000).

The described methods are currently being applied to the large HES data base of digital spectra, in order to select interesting stellar objects in an automated fashion, and fully exploit the large scientific potential present in our data base.

Our algorithms can easily be adapted for automatic classification of other large data sets, e.g. those to be compiled by the DIVA and GAIA missions.

Acknowledgements. We thank the other members of the "Methods of Scientific Discovery" group – A. Nelke, and A. Schlemminger – for their valuable contributions. The grid of synthetic spectra was kindly provided by J. Reetz. Discussions with C. Hennig on mathematical aspects of our work are acknowledged. We are grateful to C. Bailer-Jones for comments on an earlier version of this paper. This work was supported by Deutsche Forschungsgemeinschaft under grants Re 353/40 and Gr 968/3.

References

Aitchison, J., Habbema, J. D. F., & Kay, J. W. 1977, App. Stat., 26, 15

Anderson, T. 1984, An Introduction to Multivariate Statistical Analysis, 2nd ed. (New York: Wiley & Sons)

Bailer-Jones, C. 2001, in Automated Data Analysis in Astronomy, ed. R. Gupta, H. Singh, & C. Bailer-Jones (New Delhi: Narosa Publishing House) [astro-ph/0102223]

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

Beers, T. C., Kage, J. A., Preston, G. W., & Shectman, S. A. 1990, AJ, 100, 849

Beers, T. C., Rossi, S., Norris, J. E., Ryan, S. G., & Shefler, T. 1999, AJ, 117, 981

Brosch, N., Brook, A., Wisotzki, L., et al. 2000, MNRAS, 313, 641

Christlieb, N. 2000, Ph.D. Thesis, University of Hamburg, http://www.sub.uni-hamburg.de/disse/209/ncdiss.html

Christlieb, N. & Beers, T. C. 2000, in Subaru HDS Workshop on stars and galaxies: Decipherment of cosmic history with spectroscopy, ed. M. Takada-Hidai, & H. Ando (Tokyo: National Astronomical Observatory), 255 [astro-ph/0001378]

Christlieb, N., Graßhoff, G., Nelke, A., Schlemminger, A., & Wisotzki, L. 1998, in Astronomical Data Analysis and Software Systems VII, ed. R. Albrecht, R. Hook, & H. Bushouse, ASP Conf. Ser., 145, 457

Christlieb, N., Green, P., Wisotzki, L., & Reimers, D. 2001a, A&A, 375, 366

Christlieb, N., Wisotzki, L., Reimers, D., Homeier, D., Koester, D., & Heber, U. 2001b, A&A, 366, 898

Corbally, C. J., Gray, R. O., & Garrison, R. F. (eds.) 1994, The MK Process at 50 Years (San Francisco: Astronomical Society of the Pacific), ASP Conf. Ser., 60

Deichsel, G., & Trampisch, H. J. 1985, Clusteranalyse und Diskriminanzanalyse (Stuttgart: Gustav Fischer Verlag)

Hand, D. J. 1981, Discrimination and Classification (New York: Wiley & Sons)

Hennig, C., & Christlieb, N. 2002, Computational Statistics & Data Analysis, in press

Houk, N. 1975, Catalogue of two-dimensional spectral types for the HD stars, vol. 1 (University of Michigan, Ann Arbor)

Houk, N. 1978, Catalogue of two-dimensional spectral types for the HD stars, vol. 2 (University of Michigan, Ann Arbor)

Houk, N. 1982, Catalogue of two-dimensional spectral types for the HD stars, vol. 3 (University of Michigan, Ann Arbor)

Houk, N. & Smith-Moore, M. 1988, Catalogue of two-dimensional spectral types for the HD stars, vol. 4 (University of Michigan, Ann Arbor)

Ikuta, C., & Arimoto, N. 1999, PASJ, 51, 459

Kurtz, M. J. 1984, in The MK Process and Stellar Classification, ed. R. F. Garrison (Toronto: David Dunlap Observatory), 136

Lang, K. R. 1992, Astrophysical Data: Planets and Stars (New York: Springer)

LaSala, J. 1994, in The MK Process at 50 Years, ed. C. J. Corbally, R. O. Gray, & R. F. Garrison (San Francisco: Astronomical Society of the Pacific), ASP Conf. Ser., 60, 312

Oey, M. 2000, ApJ, 542, L25

Penprase, B. E. 1994, in The MK Process at 50 Years, ed. C. J. Corbally, R. O. Gray, & R. F. Garrison (San Francisco: Astronomical Society of the Pacific), ASP Conf. Ser., 60, 325

Singh, H. P., Gulati, R. K., & Gupta, R. 1998, MNRAS, 295, 312

Snider, S., Allende-Prieto, C., von Hippel, T., et al. 2001, ApJ, 562, 528

Vieira, E. F., & Ponz, J. D. 1995, A&AS, 111, 393

von Hippel, T., Storrie-Lombardi, L. J., Storrie-Lombardi, M. C., & Irwin, M. J. 1994, in The MK Process at 50 Years, ed. C. J. Corbally, R. O. Gray, & R. F. Garrison (San Francisco: Astronomical Society of the Pacific), ASP Conf. Ser., 60, 289

Weaver, W. B., & Torres-Dodgen, A. V. 1995, ApJ, 446, 300

Weaver, W. B., & Torres-Dodgen, A. V. 1997, ApJ, 487, 847

Whitney, C. A. 1983, A&AS, 51, 443

Wisotzki, L., Christlieb, N., Bade, N., et al. 2000, A&A, 358, 77

Wisotzki, L., Köhler, T., Groote, D., & Reimers, D. 1996, A&AS, 115, 227