3 Automatic Spectral Classification

The main goal of automatic classification in the HES is to identify objects of a certain class in its large data base. More formally, the problem can be stated as follows. The HES data base of digital spectra can be represented by feature vectors $\vec{x}$, consisting of a set of continuous values x_i, i.e.

$$\vec{x} = (x_1, \ldots, x_d),$$

where d is the number of features used. In Sect. 3.2 we describe how features are derived from HES spectra. We want to construct a *decision rule* which allows to assign a spectrum with feature vector $\vec{x}$ to one of the n_c classes Ω_j, $j = 1 \ldots n_c$, defined in the specific classification context. That is, we want to carry out a *supervised* classification, as opposed to *un supervised* classification, where the aim is to group objects into classes *not* defined before the classification process.

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

$$\{\vec{x}\} = (\vec{x}_1, \ldots, \vec{x}_{n_{ls}}),$$

and for which the *real* classes are known. The real 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 by 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

$$p(\vec{x}|\Omega_j)$$

can be gained. $p(\vec{x}|\Omega_j)d\vec{x}$ is the probability to observe a feature vector in the range $\vec{x} \ldots \vec{x} + d\vec{x}$ in the class Ω_j. Experience has shown that in most HES applications it is appropriate to model $p(\vec{x}|\Omega_j)$ by multivariate normal distributions, i.e.,

$$p(\vec{x}|\Omega_j) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma_j|}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_j)\,\Sigma_j^{-1}\,(\vec{x}-\vec{\mu}_j)'\right\}, \tag{6}$$

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

In many applications of automatic spectral classification in the HES, it is not possible to generate a large enough learning sample from *real* spectra present on HES plates. This is because usually the target objects are very rare. Therefore, we have developed methods to generate *artificial* learning samples by simulations, using either model spectra, or slit spectra. The next section is devoted to a description of the procedures involved.

3.1 Simulation of Objective-Prism Spectra

The conversion of model spectra, or slit spectra, to objective-prism spectra consists of 5 steps:

(1) Rebinning to the non-equidistant pixel size according to the global dispersion relation (Eq. 3)

(2) Multiplication with the HES spectral sensitivity curve(s)

(3) Smoothing with a Gaussian filter, for simulation of the seeing profile

(4) Adding of pixel-wise, normally distributed noise

(5) Random shift of the simulated spectrum according to the error distribution of the wavelength calibration zero point ($\pm 10\,\mu$m).

Step (4) ensures that objects of any brightness can be simulated; the average brightness corresponding to a given S/N can be derived from Eq. (5).

3.1.1 HES Spectral Sensitivity Curves

Spectral sensitivity curves (SSCs) for HES plates have been determined by comparison of white dwarf model spectra, rebinned to the wavelength dependent pixel size $\Delta\lambda$ of the objective-prism spectra, with objective-prism spectra of DA white dwarfs on HES plates. A first implementation of such a procedure was done by von Laar (1995). The DA model spectra were fitted to slit spectra of each of the white dwarfs under comparison. We do not use the slit spectra directly as reference, because slit losses would produce erroneous results.

By comparing SSCs for plates from different plate batches, with different sky background, and generated with objects spanning a wide brightness range (but below the saturation threshold), we investigated the possible systematic influence of these characteristics on the shape of the SSCs. The parameters of the 12 objects used in this investigation, and plate parameters, are listed in Tab. 2.

Name	B_J	Plate	bgr	Batch	$T_{\rm eff}$ [K]	$\log g$
HE 0004-5403	16.2	12076	1123	1D4	$18\,200 \pm 300$	8.26 ± 0.06
HE 0059-5701	16.4	12052	1026	1D4	$30\,400 \pm 300$	8.08 ± 0.06
HE 0252-3501	16.0	11420	1039	1D4	$17\,400 \pm 300$	7.35 ± 0.05
HE 0358-5127	15.4	10844	765	1I3	$24\,100 \pm 300$	8.10 ± 0.05
HE 0409-5154	16.1	10844	765	1I3	$27\,500 \pm 300$	8.00 ± 0.06
HE 0412-4744	16.5	10844	765	1I3	$19\,300 \pm 300$	8.08 ± 0.06
HE 0418-5326	16.1	10939	649	1I3	$27\,900 \pm 200$	8.00 ± 0.05
HE 1049-1552	14.2	9091	752	1C8	$20\,200 \pm 200$	8.63 ± 0.04
HE 1058-1258	14.8	9091	752	1C8	$24\,700 \pm 200$	8.84 ± 0.04
HE 1058-1334	16.6	9091	752	1C8	$15\,900 \pm 300$	8.00 ± 0.07
HE 1017-1618	15.8	8402	1363	1K6	$28\,600 \pm 300$	8.30 ± 0.06
HE 1017-1352	14.4	8402	1363	1K6	$33\,500 \pm 200$	8.25 ± 0.05

Table 2: Sample of DA white dwarfs used for determination of spectral sensitivity curves. bgr is the diffuse background (in counts) averaged over four plate quarters.

By comparing the shapes of the 12 resulting SSCs, we found that there is *no* systematic influence of object brightness, plate batch and sky background on SSC shape. The plate material of the HES is astonishingly homogenous as compared to other surveys, e.g. the other Hamburg based objective prism survey, the HQS Hagen et al. (1995). However, a *slight* variation of SSC shape is present (see Fig. 9), which hence must be attributed to another parameter. Since it is the blue part of the SSCs that varies, it is very likely that the time span between hypersensitization and development of the plate is responsible for the shape variations.

We grouped the 12 SSCs into 4 SSC classes of similar shape (see Tab. 3), and averaged them within these classes. When converting model spectra or slit spectra to objective prism spectra, we use an SSC created by averaging the 4 averaged SSCs with randomly assigned weights.

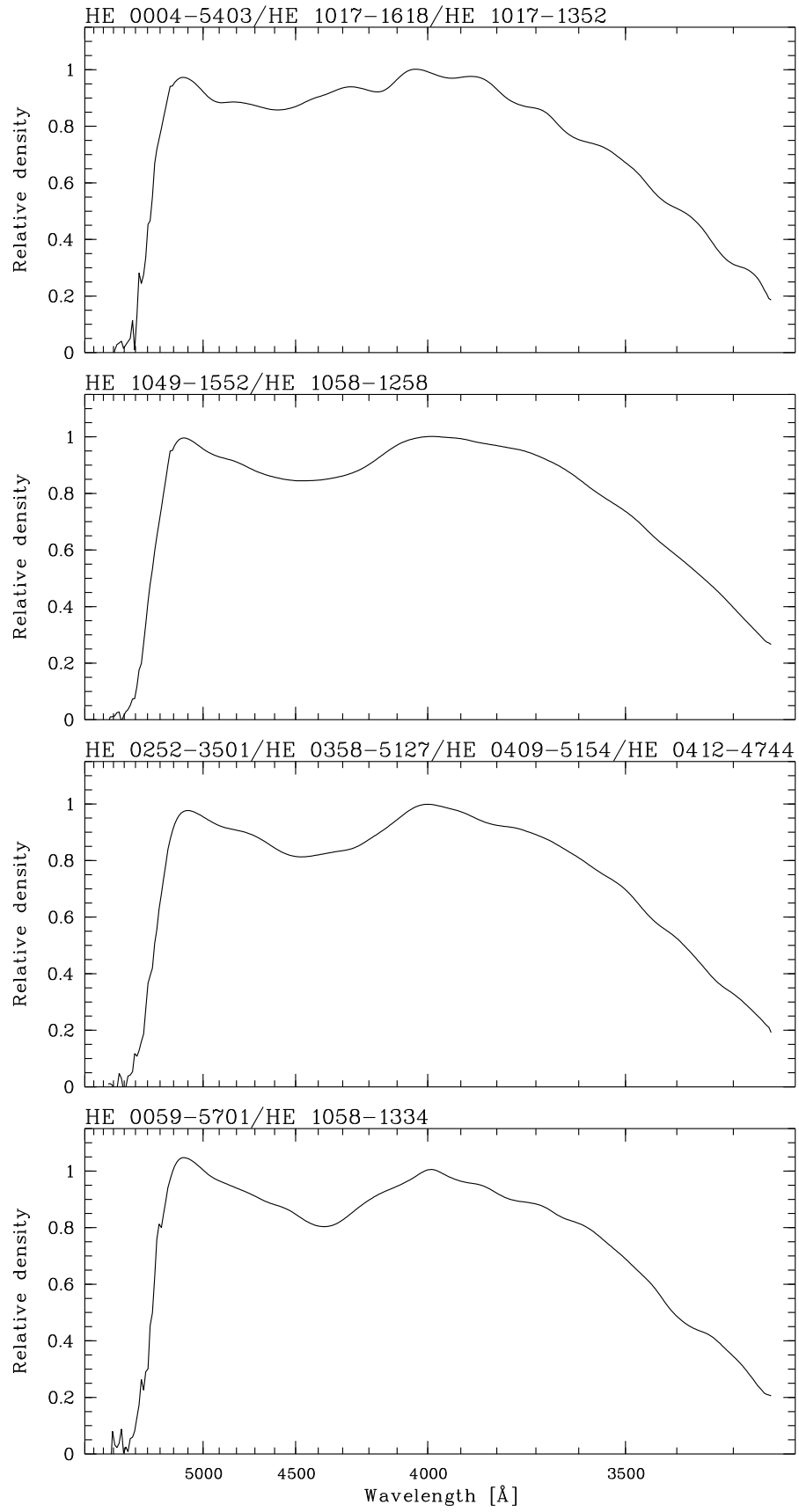


Figure 9: Averaged spectral sensitivity curves.

#	Name	B_J	Plate	bgr	Batch
1	HE 0004-5403	16.2	12076	1123	1D4
1	HE 1017-1618	15.8	8402	1363	1K6
1	HE 1017-1352	14.4	8402	1363	1K6
2	HE 1049-1552	14.2	9091	752	1C8
2	HE 1058-1258	14.8	9091	752	1C8
3	HE 0252-3501	16.0	11420	1039	1D4
3	HE 0358-5127	15.4	10844	765	1I3
3	HE 0409-5154	16.1	10844	765	1I3
3	HE 0412-4744	16.5	10844	765	1I3
4	HE 0059-5701	16.4	12052	1026	1D4
4	HE 1058-1334	16.6	9091	752	1C8

Table 3: Averaging of spectral sensitivity curves of similar shape.

3.1.2 Adding Noise

We add artificial, normally distributed noise to the converted spectra, in order to simulate objective-prism spectra of any brightness. We parameterize the S/N of a spectrum by the mean S/N in the B_J band,

$$\overline{\left(\frac{S}{N}\right)}_{B_J} = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i}{a_0 + a_1 D_i + a_2 D_i^2},$$

using the noise model described in Sect. 2.2.6. Since the noise depends on the density D, it is important to take care of the density variation throughout the spectrum. We thus scale the simulated spectra with a scaling factor c such that the desired mean S/N in B_J is achieved, when the appropriate amount of pixel-wise Gaussian noise is added. We use the typical noise coefficients

$$a_0 = 18.4 \quad (7)$$
$$a_1 = 0.604 \cdot 10^{-2} \quad (8)$$
$$a_2 = 0.719 \cdot 10^{-5}. \quad (9)$$

The mean S/N of the scaled spectrum is:

$$\overline{\left(\frac{S}{N}\right)}_{\text{new}} = \frac{1}{n}\sum_{i=1}^{n}\frac{c \cdot D_i}{a_0 + a_1 \cdot c \cdot D_i + a_2 \cdot c^2 \cdot D_i^2}, \quad (10)$$

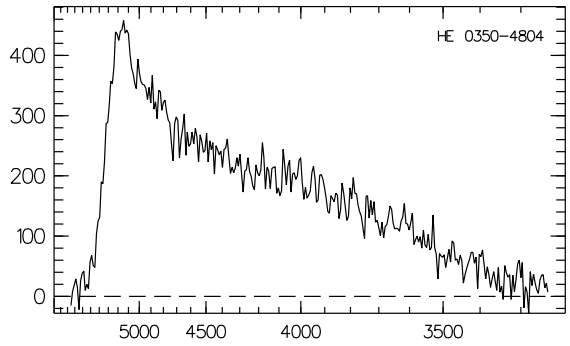
For the determination of c we re-arrange this equation to:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{c \cdot D_i}{a_0 + a_1 \cdot c \cdot D_i + a_2 \cdot c^2 \cdot D_i^2} - \overline{\left(\frac{S}{N}\right)}_{\text{new}} = 0 \quad (11)$$

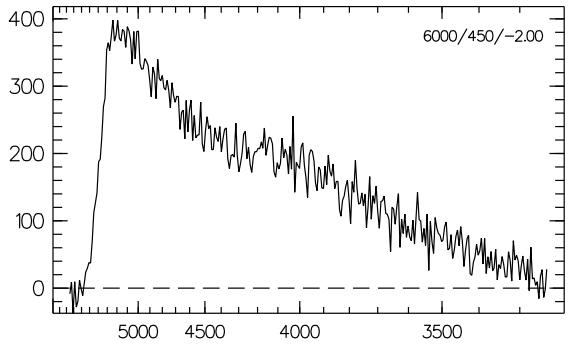
Eq. (11) can be solved iteratively with the Newton-Raphson method. The iteration rule is:

$$c^{(m+1)} = c^{(m)} - \frac{f(c^{(m)})}{f'(c^{(m)})}, \quad (12)$$

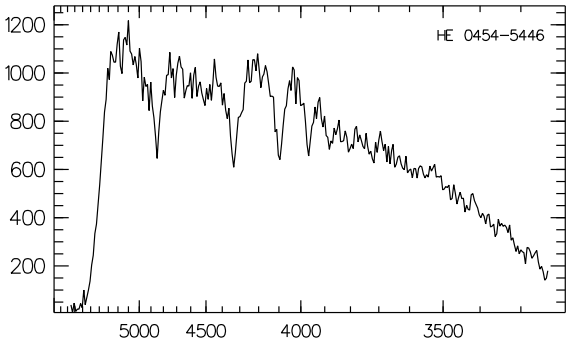
where $f(c)$ is the left hand side of Eq. (11). A comparison of simulated spectra with real spectra is shown in Fig. 10.



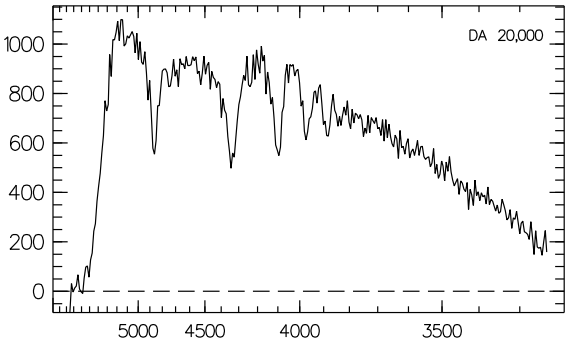
(a) Metal-poor halo star HE 0350-4804.



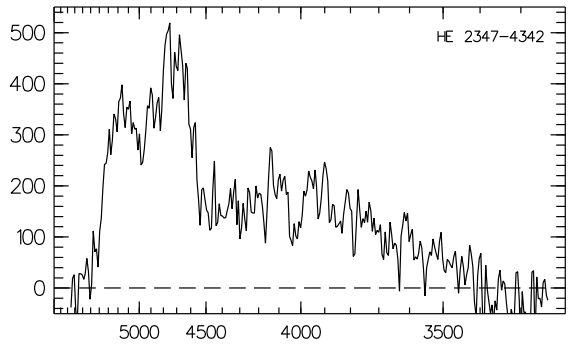
(b) Model spectrum of a metal poor star with the same stellar parameters as HE 0350-4804, converted to objective-prism spectrum and with added artificial noise.



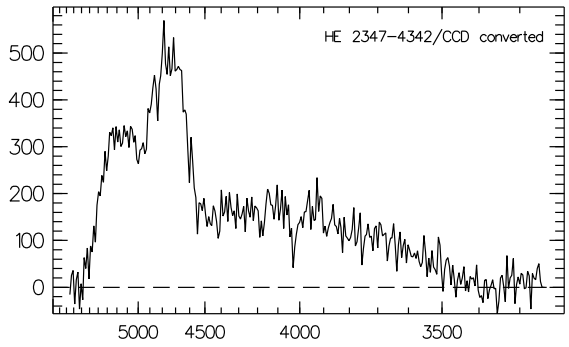
(c) HE 0454-5446, white dwarf of type DA.



(d) Model spectrum of a DA with $T_{\rm eff} = 20\,000$ K, converted to objective-prism spectrum and with added artificial noise.



(e) Quasar HE 2347-4342, $z = 2.89$.



(f) Slit spectrum of HE 2347-4342, converted to objective-prism spectrum and with added artificial noise.

Figure 10: HES spectra of objects of different type (left column) in comparison with simulated objective-prism spectra (right column). The units of the ordinates are densities in arbitrary units.

Since a lower S/N results in a larger scatter of feature values x_i within each class, the parameters of the class-conditional probabilities Eq. (6) are determined independently for artificial learning samples of different S/N. We use the following S/N grid:

$$\overline{\left(\frac{S}{N}\right)}_{B_J} = 5, 10, 15, 20, 25, 30. \tag{13}$$

Each spectrum is then classified by using the learning sample with a S/N which is closest to the S/N of the spectrum; e.g. a spectrum with $S/N = 18$ is classified by using the learning sample with $S/N = 20$.

3.2 Feature Detection

It is critical for automatic classification to have a set of *reliable* features at hand. The total set of available features should contain as much information of the objects to be classified as possible.

A wide range of spectral features is automatically detected from the digitized objective-prism spectra during the data reduction process (see Tab. 4): stellar absorption and emission lines, continuum shape, Calcium-break, bisecting points of spectral density distribution, C_2 and CN band indices, and a Ca K line index.

Name	Description	Detection method
`all5160eqw`	W_λ of Mg I b triplett/TiO λ 5168	Iterative fit procedure
`all4861eqw`	W_λ of Hβ	Iterative fit procedure
`all4388eqw`	W_λ of Fe I λ 4383+85	Iterative fit procedure
`all4340eqw`	W_λ of Hγ	Iterative fit procedure
`all4300eqw`	W_λ of G-Band	Iterative fit procedure
`all4261eqw`	W_λ of Cr I λ 4254 + 75 + Fe I 4260 + 72	Iterative fit procedure
`all4227eqw`	W_λ of Ca I λ 4227	Iterative fit procedure
`all4102eqw`	W_λ of Hδ	Iterative fit procedure
`all3969eqw`	W_λ of Ca H + Hε	Iterative fit procedure
`all3934eqw`	W_λ of Ca K	Iterative fit procedure
`klcomp_1`	1. continuum shape coefficient	PCA
`klcomp_2`	2. continuum shape coefficient	PCA
`klcomp_3`	3. continuum shape coefficient	PCA
`klcomp_4`	4. continuum shape coefficient	PCA
`CaBreak_sn`	S/N Calcium-break	Template matching
`CaBreak_cont`	Contrast of Calcium-break to continuum	Template matching
`CaKindex`	Strength of Ca K	Ratio of average pixel values
`C2idx1`	Strength of C_2 λ 5165	Ratio of average pixel values
`C2idx2`	Strength of C_2 λ 4737	Ratio of average pixel values
`CNidx2`	Strength of CN λ 4216	Ratio of average pixel values
`CNidx3`	Strength of CN λ 3883	Ratio of average pixel values
`dx_hpp1`	Half power point distance 1	Summing of pixel values
`dx_hpp2`	Half power point distance 2	Summing of pixel values

Table 4: Automatically detected spectral features in the HES.

3.2.1 Detection of Stellar Lines

We implemented a flexible, robust algorithm which allows to detect stellar absorption and/or emission lines in HES spectra. The algorithm is iterative, consisting of the following steps:

(1) Determination of continuum by filtering with a wide median filter and narrow Gaussian filter.

(2) Improvement of determination of the wavelength calibration zero point by fitting of 3 sets stellar lines. The sets contain the strongest stellar absorption lines of early type, solar type, and late type stars, respectively. The individual lines depths, and the zero point offset of wavelength calibration are fitted simultaneously. The *relative* positions of the stellar lines are held fixed, and the line *widths* is held fixed at the value of the seeing profile widths, which is measured during spectral extraction. The set of lines giving the strongest signal, i.e. largest average equivalent widths, is selected, and the wavelength calibration zero point determined with that fit is adopted.

(3) Improvement of continuum determination:

  (a) Fitting of *all* stellar lines detectable in HES spectra

  (b) Removal of fitted lines from the original spectrum

  (c) Computation of improved continuum by filtering the line-reduced spectrum again with a wide median filter and narrow Gaussian filter

  (c) Start with (3a), if $n_{\rm iter} < 3$; otherwise compute rectified spectrum with final continuum.

(4) Fitting of all stellar lines in the rectified spectrum by Gaussians.

For each spectral line it can be chosen whether it is to be detected in absorption or emission. The output of the fit algorithm are equivalent width, FWHM and S/N of the lines, and shift of the wavelength calibration zero point. Any spectral lines not yet considered can easily be included by just adding its wavelength to the list of lines to be fitted.

3.2.2 PCA of Continua

We perform a Principal Component Analysis (PCA; see e.g. Murtagh & Heck 1987) of the continua determined in the iterative line detection procedure, in order to parameterize the continuum shape of HES spectra. Wisotzki (1991) used PCA in the HES also for quasar selection.

It was found that for the continua of a learning sample of 654 spectra classified by hand, and occupying the classes A5–K9, three principal components account for almost 98 % of the variance in the learning sample. It was possible to fit the continua of almost all of the learning sample spectra with $\chi^2/\nu < 2.0$. Four objects needed 5 or more components. Their spectra were inspected again and it was found that they have an unusual continuum shape, probably because they are binaries. These spectra have thus been excluded from the learning sample.

3.2.3 Broad Band Colours

For many stellar applications it is helpful to have $U-B$ and $B-V$ colours at hand. Therefore, we established colour calibrations using so-called "half power points" (hpp; see Wisotzki et al. 2000). These are bisecting points of a part of the spectrum. Definitions of the hpps can be found in Tab. 5, and an illustration in Fig. 12. hpps are equivalent to broad band colours, but have the advantage of being more robust against noise. `x_hpp1` and `x_hpp2` are well correlated with $U-B$ and $B-V$, respectively.

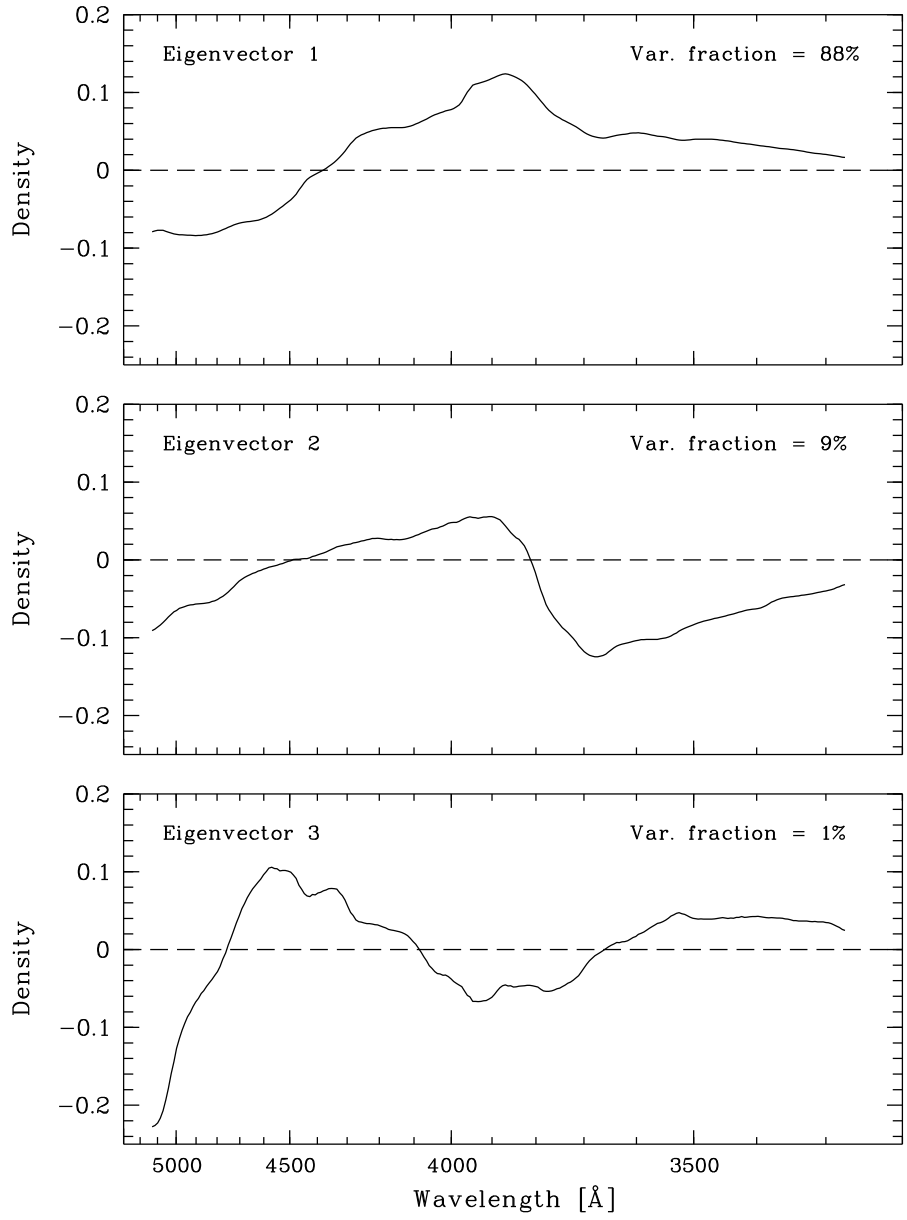


Figure 11: First three principal components of the continua of 654 learning sample spectra.

Name	λ range	correlated with
x_hpp1	$3240\,\text{Å} < \lambda < 4840\,\text{Å}$	$U-B$
x_hpp2	$3890\,\text{Å} < \lambda < 5360\,\text{Å}$	$B-V$

Table 5: Definitions of spectral half power points (hpp) used in the HES.

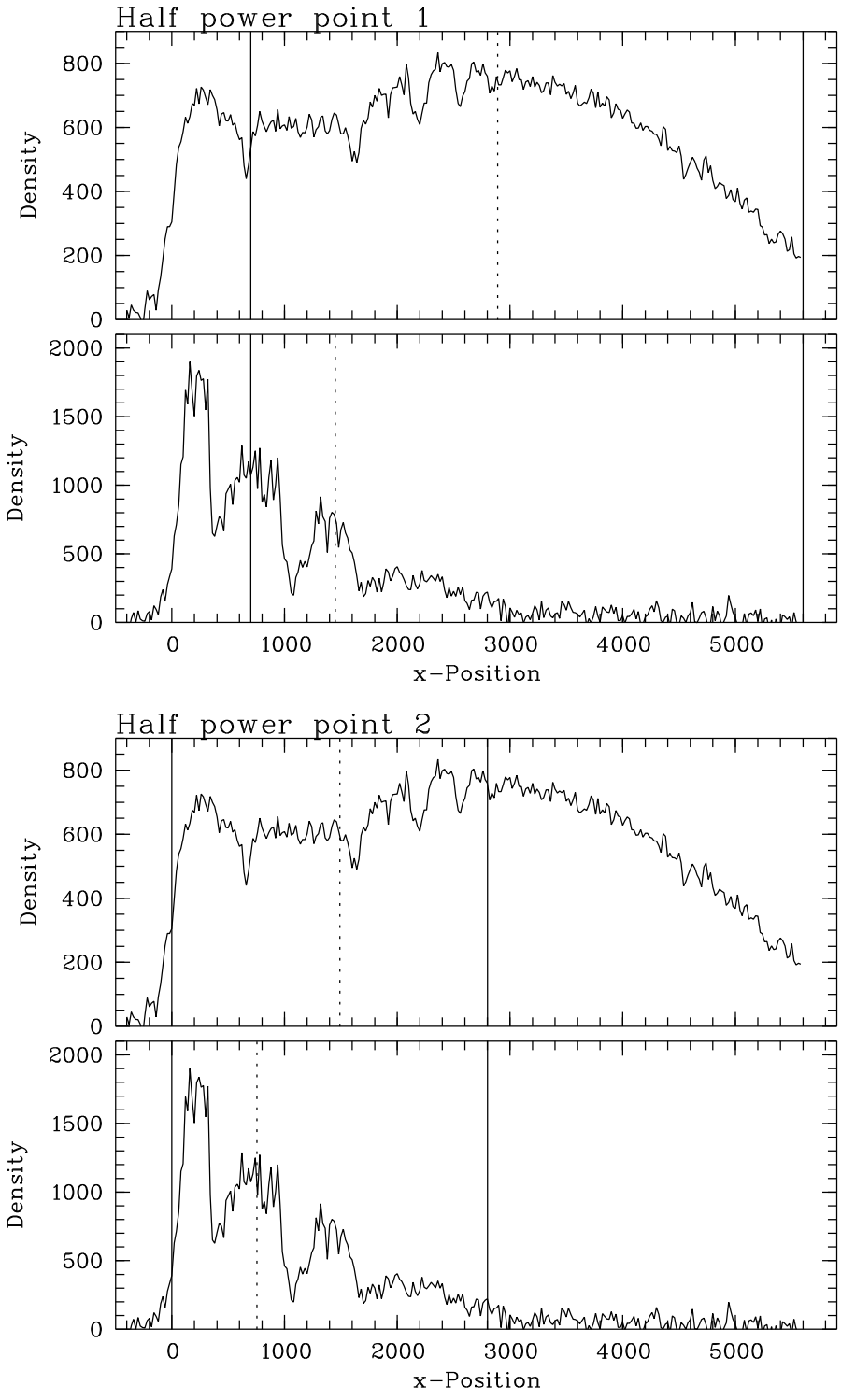


Figure 12: Illustration of spectral half power points `x_hpp1` and `x_hpp2`. Solid lines mark the regions in which the hpps are computed; dotted lines indicate the position of the hpps.

A more precise colour calibration can be achieved when distances `dx` to a cutoff line in a colour-magnitude diagram (see Fig. 13) is used instead of `x` values for the bisecting point, because plate-to-plate variations of the spectral sensitivity curves are compensated in this way. The cutoff line separates the bulk of "normal" stars from UV-excess objects (or objects with unusually low $B-V$ in case of `dx_hpp2`). The cutoff is determined by a break finding algorithm.

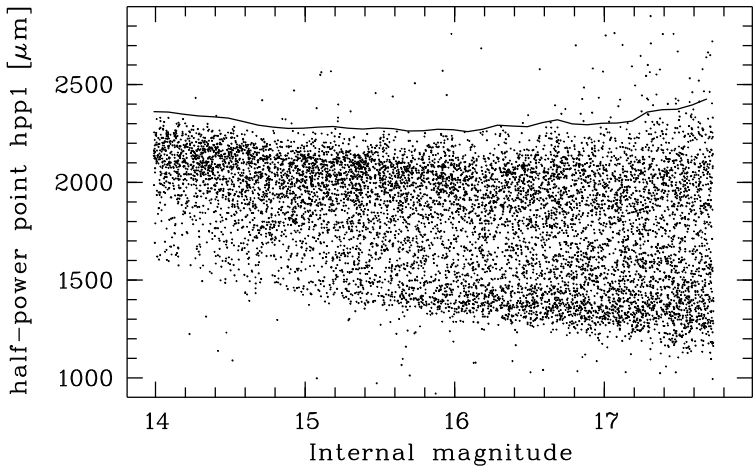


Figure 13: Cutoff-line for bisecting point `x_hpp1` on one HES plate.

Because the blue end of the HES spectra is sensitive to contamination by overlaps, special care must be taken to exclude such spectra from the calibration of `dx_hpp1`. This has been done by applying stricter overlap selection criteria. In addition, an iterative $\kappa\sigma$-clipping with $\kappa = 3$ has been employed to exclude overlaps unrecognized by the automatic detection. 50 of the 623 spectra in the original data set have been clipped out, so that the calibration uses spectra of 573 objects. The resulting fit is shown in Fig. 14.

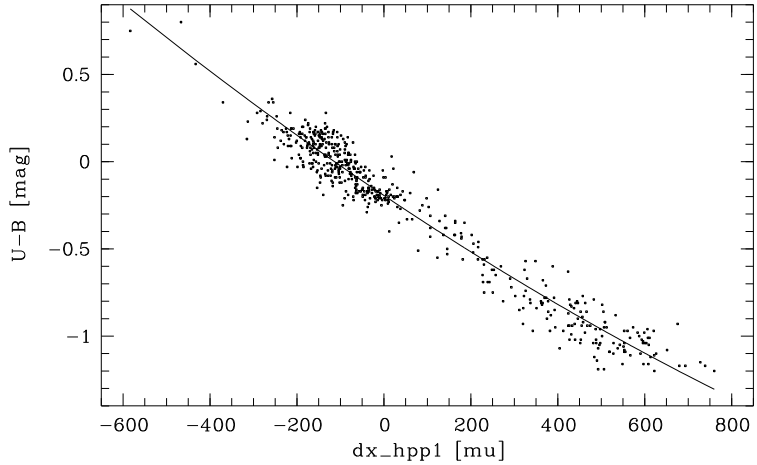


Figure 14: Calibration of `dx_hpp1` using a combined sample of 573 objects from the EC and HK surveys present on HES plates.

A potential problem for the $B-V$ calibration is that the V band is not fully covered by the HES wavelength range. Therefore, the calibration for very red objects is inaccurate, or even impossible. As calibrators for red objects, 36 carbon stars have been used, for which BV photometry was obtained by

the author at the ESO 2.2 m telescope in April 1999. Carbon stars with $B-V > 2.5$ have been excluded from the fit. For $B-V \lesssim 1.0$, 778 stars from the HK survey of Beers et al. (1992), 354 FHB and other A-type stars of Wilhelm et al. (1999), and 272 objects from the northern galactic cap fields of the EC survey (Kilkenny et al. 1997) present on HES plates have been used. Linear fits in three colour regions have been done separately, in order to evaluate the scatter independently, and check consistency. Then, a combined fit to all 1256 unique objects was done (see Fig. 15).

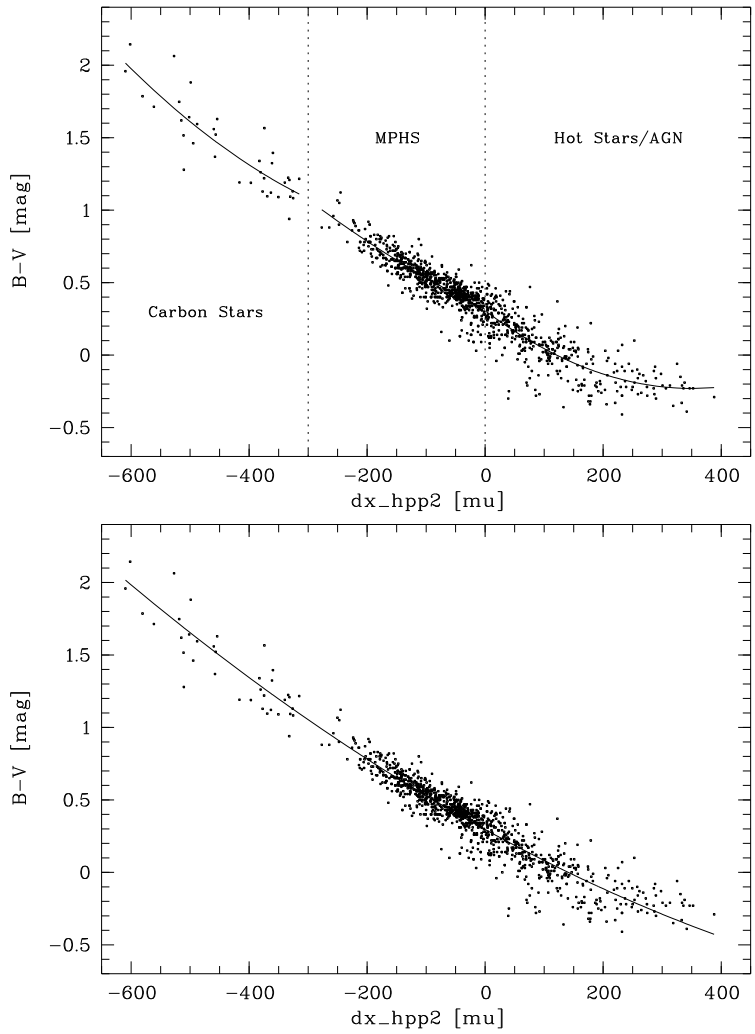


Figure 15: Calibration of dx_hpp2 by separate fits in different colour regions (upper panel), and by combined fit to 1256 objects (lower panel).

The results of the fits are summarized in Tab. 6. Note that a single fit contains objects from a large fraction of the 329 stellar HES plates, and – with the exception of the red $B-V$ fit – a wide range of object types, e.g. metal-poor stars, solar metallicity F- and G-type stars, field horizontal branch A-type stars, "normal" A-type stars, DA white dwarfs, DB white dwarfs, sdB stars, AGN. The achieved accuracies are $\sigma_{U-B} = 0.092^m$, and $\sigma_{B-V} = 0.095^m$ for the $B-V$ fit using all calibration objects together. The accuracy in $B-V$ for red ($B-V \gtrsim 1$) and blue ($B-V \lesssim 0.3$) objects is a factor of ~ 2 worse ($\sigma = 0.15^m$ and 0.12^m, respectively) than for intermediate $B-V$ objects ($\sigma = 0.074^m$).

Colour	a_0	a_1	a_2	valid range	N_{stars}	σ [mag]
$B-V$	0.79	$2.53\cdot10^{-5}$	$3.34\cdot10^{-6}$	$-600 <$ dx_hpp2 < -300	37	0.15
$B-V$	0.31	$-2.00\cdot10^{-3}$	$1.74\cdot10^{-6}$	$-300 <$ dx_hpp2 < 0	817	0.074
$B-V$	0.31	$-3.06\cdot10^{-3}$	$4.35\cdot10^{-6}$	$0 <$ dx_hpp2 < 400	405	0.12
$B-V$	0.30	$-2.24\cdot10^{-3}$	$9.62\cdot10^{-7}$	$-600 <$ dx_hpp2 < 400	1259	0.095
$U-B$	-0.19	$-1.67\cdot10^{-3}$	$+2.76\cdot10^{-7}$	$-800 <$ dx_hpp1 $< +800$	573	0.092

Table 6: Broad band colour calibration fits.

3.2.4 Narrow Band Colours

We obtain Strömgren coefficients $c_1 = (u-b)-(v-b)$ directly from HES spectra by averaging the density in the Strömgren *uvb* bands, and computing internal coefficients $c_{1,\mathrm{HES}}$ from that. $c_{1,\mathrm{HES}}$ has been calibrated using a total of 79 stars, which are not saturated in the HES, from three different sources. 22 metal-poor stars have been taken from Schuster et al. (1996), 43 stars from Beers (2000, priv. comm.), of which 2 have been rejected as outliers (see Fig. 16), and 16 hot subdwarfs from an updated version of the catalog of Kilkenny et al. (1988) (Heber 2000, priv. comm.). The 1σ error of the calibration is 0.15^m. c_1 can be used as a gravity indicator, since it measures the strength of the Balmer discontinuity.

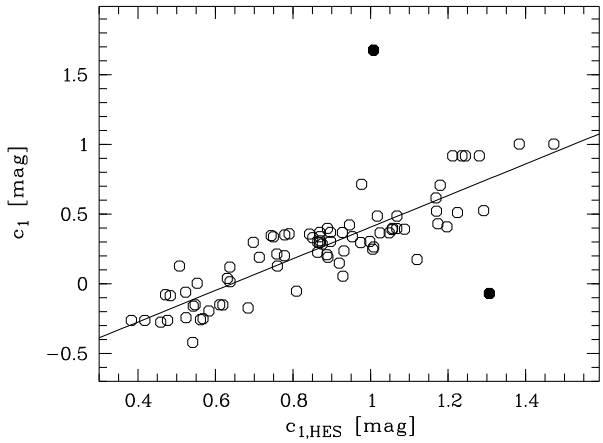


Figure 16: Calibration of Strömgren c_1 measured in HES spectra. The 2 filled circles mark objects excluded from the fit.

3.3 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 two decision rules: The Bayes rule, and a minimum cost rule.

3.3.1 Bayes' Rule

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

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

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

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

3.3.2 Minimum Cost Rule

In most of the classification problems arising in the HES it is desired to compile 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 compile 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 compiled, 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

$$0 \le r_{hk} \le 1; \qquad h = 1,\ldots,n_c; \quad k = 1,\ldots,n_c. \tag{14}$$

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 $\vec{x}$. 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 $C_{h\to k}(\vec{x})$ is:

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

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

$$C_k(\vec{x}) = \sum_{\substack{h=1 \\ h\neq k}}^{m} C_{h\to k}(\vec{x})$$

$$= \sum_{\substack{h=1 \\ h\neq k}}^{m} r_{hk} \frac{a_h p_h(\vec{x})}{\sum_{i=1}^{m} a_i p_i(\vec{x})} \tag{15}$$

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

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

If the cost factors have been chosen such that

$$r_{hk} \equiv \delta_{hk},$$

the minimum cost rule classification is identical to classification according to Bayes' rule. In this case the cost for assigning the class Ω_k to a spectrum with feature vector $\vec{x}$ is the probability that the object belongs to one of the other classes $h \neq k$. This follows immediately from Eq. (15). 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.

3.3.3 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 $\vec{x}$ may be assigned to a class with *very low* posterior probability $p(\Omega_i|\vec{x})$, if $p(\Omega_i|\vec{x})$ is even lower for all other classes. This means that a class is assigned to *all* spectra, even to "garbage spectra" which have been disturbed, for instance, by plate artifacts. Therefore, it is useful to make use of a rejection criterion.

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

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

$$a.i.(\Omega_i,\vec{x}) = \Gamma\left\{\frac{d}{2};\frac{1}{2}(\vec{x}-\vec{\mu}_i)\,\Sigma_i^{-1}\,(\vec{x}-\vec{\mu}_i)'\right\},$$

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 $\vec{x}$ 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 *un*likely that a feature vector $\vec{x}$ is observed for class Ω_i, given the multivariate normal distributions (6) are the *true* distributions of the class-conditional probabilities $p(\vec{x}|\Omega_i)$.

3.4 Evaluation of Classification Rules

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

(1) Re-substitution

(2) "Hould 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 misclassification 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 disjunctly 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. When using non-parametric methods, the situation is even worse, because the *shape* of the distribution has to be determined, too, so that additional degrees of freedom are present.

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_{ls}. We exclude object i from the learning sample, and construct the classification rule using the $n_{ls} - 1$ remaining objects. Object i is then classified with this classification rule. This procedure is repeated n_{ls} 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_{ls} classification rules have to constructed. However, the computing time increases only *linearily* with learning sample size n_{ls}, 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 $n_{ls} = 165\,000$).

3.5 Choosing a Feature Combination

It is necessary to select a subset of the available features for each classification problem, and each S/N step, 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 less features results in more stable estimates of the parameters of the multivariate normal distributions Eq. (6).

(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 colours for classification, because no stellar lines can be detected reliably anymore.

The best method for finding the optimal feature combination is to evaluate *all* $2^d - 1$ possible combinations of the d available features, since this is the only way to *prove* that the combination found is really the best one. However, since the computing time raises exponentially with the number of features, the complete search is only feasible for a limited number of features. On a Linux PC

with 333 MHz Pentium II processor the complete search in a feature space of $d = 11$, evaluated with "leaving one out" on a learning sample of 22 500 objects, takes about one day. Since the search has to be done for all 6 S/N steps individually, $d = 11$ is about the feasibility limit.

In practice it is usually possible to select a subset of $d < 11$ features from the 23 available features listed in Tab. 4 by astrophysical considerations alone. E.g., when it is desired to select metal-poor stars, one can restrict the initial feature set to those features that are *possibly* useful as indicators for $T_{\rm eff}$, $\log g$, and [Fe/H], and one can safely ignore e.g. Carbon band indices. 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.

3.6 Choosing Cost Factors

The cost factors were adjusted by using a special tool, which displays the confusion matrix, estimated with the "leaving one out" method on the learning sample, depending on the choice of three sets of cost factors:

`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. A screen-dump of the cost factor tool is shown in Fig. 17.

Acknowledgements

The work described in this section was partly carried out in collaboration with the research group *Epistemische Systeme* lead by G. Graßhoff. Together with A. Nelke and A. Schlemminger, we critically discussed methods of automatic classification we found in the literature, and we developed many ideas during the legendary Saturday morning meetings, which never lasted less than 4 hours! A. Nelke has also implemented the cost factor adjusting tool.

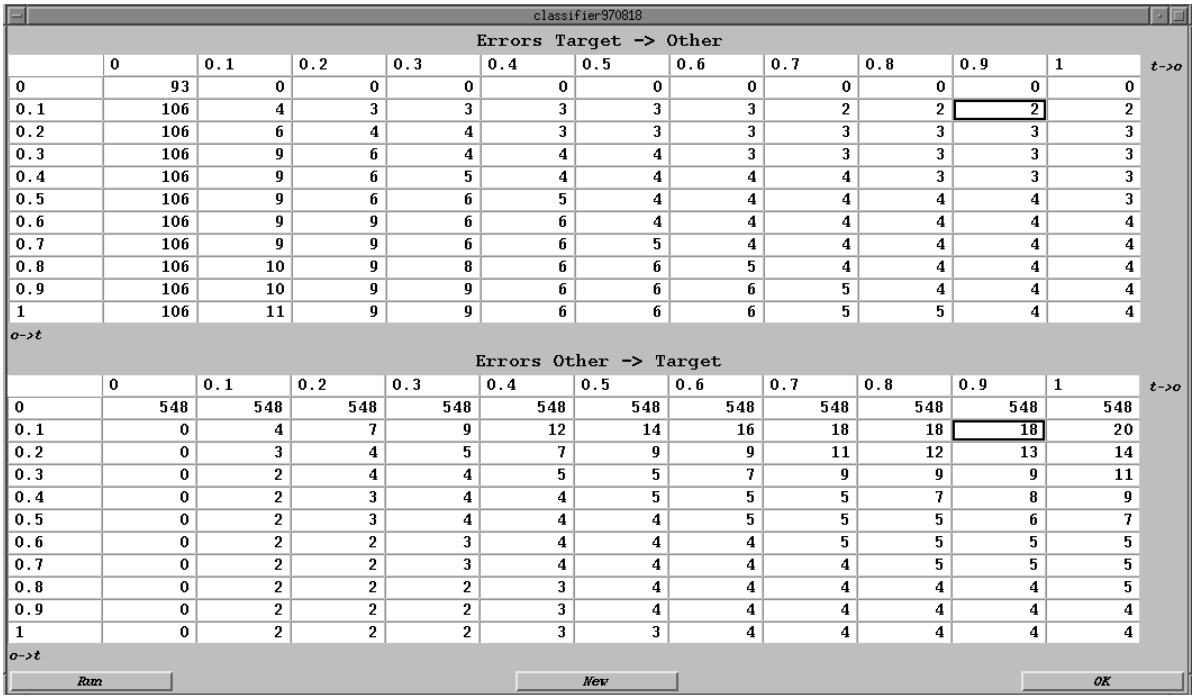
I thank D. Homeier for fitting DA models to slit spectra, and V. Beckmann for obtaining a part of those at ESO. G. Pizarro kindly compiled a list of HES plate batches from the notes he made at the ESO Schmidt telescope.

Precise, photoelectric *UBV* and Strömgren *uvby* photometry for HK survey stars was kindly provided by T. Beers before publication. I thank D. O'Donoghue for making EC survey photometry in digital form available to me. U. Heber provided an updated version of the subdwarf catalog of Kilkenny et al. (1997).

References

Aitchison, J., Habbema, J. & Kay, J. (1977), 'A Critical Comparison of Two Methods of Statistical Discrimination', *Applied Statistics* **26**(1), 15–25.

Anderson, T. (1984), *An Introduction to Multivariate Statistical Analysis*, 2 edn, Wiley & Sons, New York.



	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	t->o
0	93	0	0	0	0	0	0	0	0	0	0	
0.1	106	4	3	3	3	3	3	2	2	2	2	
0.2	106	6	4	4	3	3	3	3	3	3	3	
0.3	106	9	6	4	4	4	3	3	3	3	3	
0.4	106	9	6	5	4	4	4	4	3	3	3	
0.5	106	9	6	6	5	4	4	4	4	4	3	
0.6	106	9	9	6	6	4	4	4	4	4	4	
0.7	106	9	9	6	6	5	4	4	4	4	4	
0.8	106	10	9	8	6	6	5	4	4	4	4	
0.9	106	10	9	9	6	6	6	5	4	4	4	
1	106	11	9	9	6	6	6	5	5	4	4	
o->t												

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	t->o
0	548	548	548	548	548	548	548	548	548	548	548	
0.1	0	4	7	9	12	14	16	18	18	18	20	
0.2	0	3	4	5	7	9	9	11	12	13	14	
0.3	0	2	4	4	5	5	7	9	9	9	11	
0.4	0	2	3	4	4	5	5	5	7	8	9	
0.5	0	2	3	4	4	4	5	5	5	6	7	
0.6	0	2	2	3	4	4	4	5	5	5	5	
0.7	0	2	2	3	4	4	4	4	5	5	5	
0.8	0	2	2	2	3	4	4	4	4	4	5	
0.9	0	2	2	2	3	4	4	4	4	4	4	
1	0	2	2	2	3	3	4	4	4	4	4	
o->t												

Figure 17: Screen-dump of cost factor adjustment tool. It displays the number of misclassifications from the target class(es) to other classes (upper panel) and vice versa (lower panel) in dependence of the cost factors `o2t` (abscissa) and `t2to` (ordinate). By clicking on one of the matrix fields, one can zoom in to a smaller range of cost factors around the values of that field.

Beers, T. C., Preston, G. W. & Shectman, S. A. (1992), 'A search for stars of very low metal abundance. II.', *AJ* **103**(6), 1987–2034.

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

Hagen, H.-J., Groote, D., Engels, D. & Reimers, D. (1995), 'The Hamburg Quasar Survey. I. Schmidt observations and plate digitizations', *A&AS* **111**, 195–203.

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

Kilkenny, D., Heber, U. & Drilling, J. S. (1988), 'A catalogue of spectroscopically identified hot subdwarf stars', *SAAO Circ.* **12**, 1–80.

Kilkenny, D., O'Donoghue, D., Koen, C., Stobie, R. S. & Chen, A. (1997), 'The Edinburgh-Cape Blue Object Survey – II. Zone 1 – The North Galactic Cap', *MNRAS* **287**, 867–893.

Murtagh, F. & Heck, A., eds (1987), *Multivariate Data Analysis*, Reidel, Dordrecht.

Schuster, W. J., Nissen, P. E., Parrao, L., Beers, T. C. & Overgaard, L. (1996), 'uvby-β photometry of high-velocity and metal-poor stars. VIII. Stars of very low metal abundance.', *A&AS* **117**, 317–334.

von Laar, H. (1995), Spektrophotometrische Eichung von Objektivprismenplatten, Diplomarbeit, Universität Hamburg.

Wilhelm, R., Beers, T. C., Sommer-Larsen, J., Pier, J. R., Layden, A. C., Flynn, C., Rossi, S. & Christensen, P. R. (1999), 'Spectroscopy of Hot Stars in the Galactic Halo. III. Analysis of a Large Sample of Field Horizontal-Branch and Other A-Type Stars', *AJ* **117**, 2329–2380.

Wisotzki, L. (1991), Ein statistisches Verfahren zur automatischen Quasarsuche, PhD thesis, Universit"at Hamburg.

Wisotzki, L., Christlieb, N., Bade, N., Beckmann, V., Köhler, T., Vanelle, C. & Reimers, D. (2000), 'The Hamburg/ESO survey for bright QSOs. III. A large flux-limited sample of QSOs', A&A, in press.