2 The Hamburg/ESO survey

2.1 Description of Plate Material

The HES is based on IIIa-J plates taken with the ESO Schmidt telescope and its 4° prism (Wisotzki et al. 1996, 2000). It covers the magnitude range $12.0 \gtrsim B_J \gtrsim 17.5$. The magnitude limit depends on plate quality. Note that the value given for the faint limit is the completeness limit for quasar search, which we define to be the average amplitude of the pixel-wise noise in the B_J band $> 5\sigma$ above the diffuse plate background, where σ is the background noise. The *detection limit* of the HES is $B_J \simeq 18$. For stellar applications, the survey magnitude range depends on the object type searched for. In our search for metal-poor stars, we only use spectra below a saturation threshold (which roughly corresponds to $B_J > 14.0$), and spectra with $S/N > 10$. However, for most object types we adopt the 5σ magnitude limit.

The atmospheric cutoff at the blue end and the sharp sensitivity cutoff of the IIIa-J emulsion ("red edge") result in a wavelength range of $3200\,\text{Å} < \lambda < 5300\,\text{Å}$ (see Fig. 4). The spectral resolution of the HES is primarily seeing-limited, and typically 15 Å FWHM at Hγ. For plates taken during good seeing conditions, the pixel spacings chosen in the digitization process results in an under-sampling, so that in these cases the spectral resolution is also limited by the sampling.

The definition of the HES survey area makes use of the mean star density Σ and average column density of neutral hydrogen, N_H for each ESO/SERC field. In 0th order, the HES area is defined by all fields which fulfill all of the following conditions:

$$\begin{array}{rcl} +2.5^\circ > & \delta & > -78^\circ \\ \Sigma & < & 100\ \text{stars}/\text{deg}^2 \\ N_H & < & 10^{21}\,\text{cm}^{-2}. \end{array}$$

N_H has been taken from Dickey & Lockman (1990). The above conditions correspond roughly to galactic latitudes $|b| > 30\,\text{deg}$. A few fields have been added in order to have a "smooth" survey border. In result, the survey area consists of 380 fields. Objective-prism plates have been taken for all of these, and the plates have been digitized and reduced at Hamburger Sternwarte.[2] As one plate covers approximately $5 \times 5\,\text{deg}$ on the sky, the nominal survey area is $9\,500\,\text{deg}^2$, or the total southern extragalactic sky.

Of the 380 HES plates, 54 have been excluded from exploitation of the stellar content as of now. This has several reasons. First of all, at the time of preparing the plate data for processing with candidate selection algorithms, ~ 40 plates had not been digitized. Other plates have been excluded because files were corrupted due to defect storage media (these plates will have to be re-reduced, or – in a few cases – even re-scanned), or because some of the data (e.g. descriptors, names or content of table columns, etc.) was inconsistent with the rest of the data. Therefore, this work is restricted to 329 plates only.

2.2 Data Reduction

An overview of the HES data flow is given in Fig. 2.

[2] This sentence represents 11 years of work! The following people were involved (in alphabetical order): V. Beckmann, M. Brachmann, N. Christlieb, D. Groote, H. Hagen, M. Ikonomou, T. Köhler, D. Kühl, B. Kuhlbrodt, A. Müller, G. Pizzaro, O. Pizzaro, D. Reimers, H.E. Schuster, C. Vanelle, L. Wisotzki.

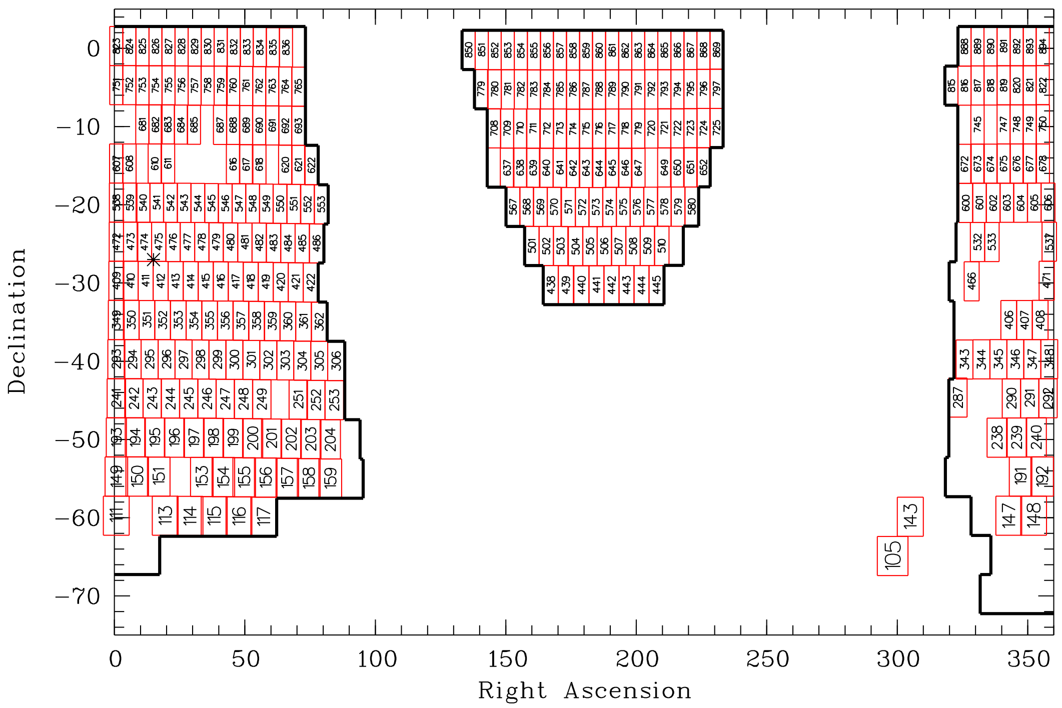

Figure 1: Definition of "official" HES area (framed) and numbers of fields in which the exploitation of the stellar content of the HES is currently carried out. For orientation, the position of the southern galactic pole is marked with '$*$' (lower left corner of field 475).

2.2.1 Plate Digitization

All HES plates have been digitized at Hamburger Sternwarte using a PDS 1010G microdensitometer (Hagen 1987). As the Hamburg PDS is too small for the large (28×28 cm) ESO plates, they have been scanned in four quadrants. The scan is performed with an $30 \times 30\,\mu$m aperture and $20\,\mu$m step width, yielding 7500×7500 pixels per plate quarter. The plate scale of the HES plates is $67\farcs5\,\mathrm{mm}^{-1}$, so that the centers of two pixels are separated by $1\farcs35$. The dispersion direction is declination, and increasing wavelength corresponds to increasing declination.

It takes about 16 hours to scan one plate. The raw data amount to ~ 430 MB per field. It is saved temporarily on magneto-optical discs (MOs) for data reduction, and after that archived on CD-ROMs. The data reduction was carried out on a dedicated PC with a Pentium 133 MHz processor, 96 MB RAM and Linux as operating system. The relevant periphery (CD-ROM drive and writer, MO drive) is attached directly to this PC, and ~ 8 GB of harddisk space is visible for it. The reduction of direct and spectral plate data of one field takes 6–8 hours.

2.2.2 Reduction of Direct Plate Data

For the automated reduction of the objective-prism plate data digitized *direct* plates of the *Digitized Sky Survey I* (DSS-I) are used (for an illustration see Fig. 3).

After background subtraction and object detection on the direct plates, an astrometric transformation between direct and objective-prism plates is computed. This yields an input catalog of object positions on the objective-prism plates for the extraction of the spectra, and the zero point for wave-

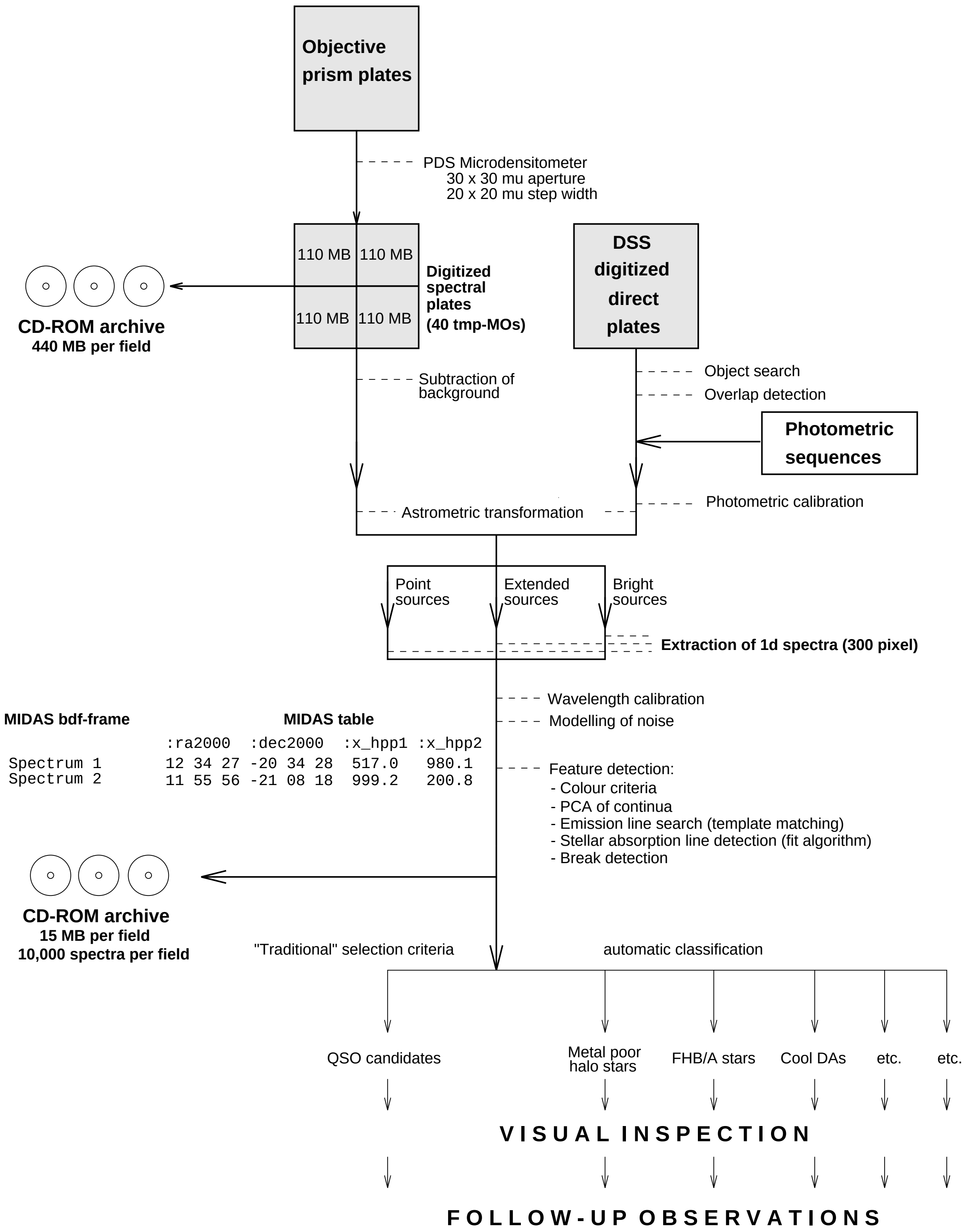

Figure 2: Overview of HES data flow.

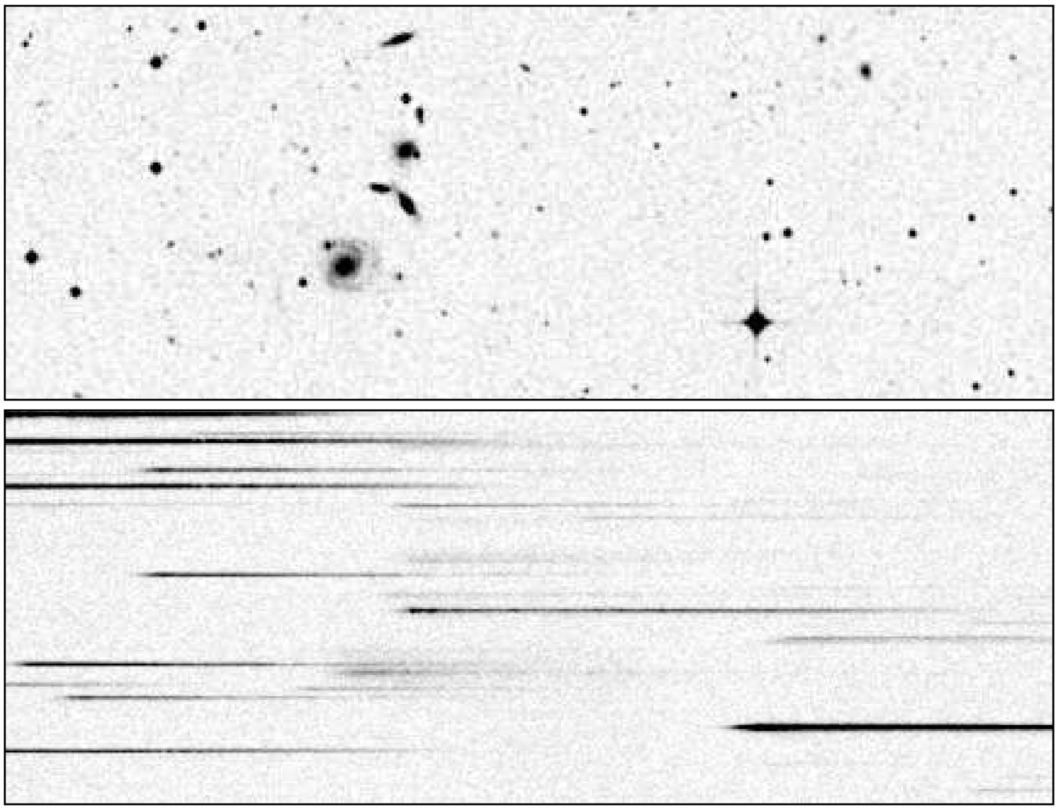

Figure 3: Direct plate scan of the *Digitized Sky Survey I* and corresponding HES plate scan. The wing-shaped object in middle height and slightly to the left in the direct scan is a pair of interacting galaxies with a Seyfert 1 nucleus (HE 0323-4204) at $z = 0.058$.

length calibration.

Although this procedure is appropriate for the primary targets of the HES, i.e. quasars, it is problematic for objects which have a large proper motion. This lesson was learned in the course of a search for dwarf carbon stars (see Sect. 5). The feature detection algorithms failed for G77-61, a dwarf carbon star with $\mu_\alpha = 0\farcs184\,\mathrm{yr}^{-1}$, $\mu_\delta = -0\farcs745\,\mathrm{yr}^{-1}$ (Deutsch 1994). Since the epoch difference Δt between the HES plate and the relevant DSS I plate is ~ 10 years (a typical value), the proper motion results in -5.6 and 1.4 pixel offsets in dispersion direction and perpendicular to the dispersion direction, respectively. Therefore, too low C_2 band index values have been measured. Proper motions along dispersion direction (i.e., declination) can at least partly be recovered by using special techniques, e.g. shifting templates through the spectrum. However, if $\mu_\alpha \cdot \Delta t$ is large, the spectrum of the object is not found on the HES plate, and only sky background is extracted.

Overlapping spectra (hereafter shortly called overlaps) are also detected with the help of the direct data: For each spectrum to be extracted, it is looked for objects in the dispersion direction on the direct plate. If there is one, the automatic procedure marks the corresponding spectrum, so that it can later be excluded from further processing, if this is desired. *It is* desired for stellar work, since the feature detection and automatic classification algorithms would get confused otherwise, and a lot of "garbage" would enter the candidate samples. The elimination of overlaps reduces the HES area from a nominal $9\,500\,\mathrm{deg}^2$ to an effective area of $\sim 7\,500\,\mathrm{deg}^2$. The overlap rate is strongly dependend on galactic lattitude b and varies between 10 % and 50 % within the survey area. 20 % is an average overlap rate for the total survey.

2.2.3 Photometry

Photometric calibration of HES plates is performed by calibrating the direct plates with individual photometric sequences, and after that calibrating the spectral plates with these magnitudes (Köhler 1991; Vanelle 1996). The overall accuracy including zero point error is less than ± 0.2 mag in B_J. Photometric sequences are available for all of the 383 HES plates, and the calibration has been carried out for all 329 plates on which stellar work is done so far.

The B_J band is formally defined by the spectral sensitivity curve of the Kodak IIIa-J emulsion multiplied with the filter curve of a Schott GG395 filter. B_J can be converted to B using the formula

$$B = B_J + 0.28 \cdot (B - V), \tag{1}$$

which is valid for main-sequence stars in the colour range $-0.1 < (B - V) < 1.6$ (Hewett et al. 1995).

2.2.4 Extraction of Objective-Prism Spectra

Extraction of the objective-prism spectra is done by optimal procedures, maximizing S/N. According to their appearance on the direct plates, and brightness on HES plates, three types of objects are extracted in different ways:

`stars` The spectra of point-like sources are extracted by fitting Gaussian profiles to the 2–3 pixels with the highest density values in each column perpendicular to the dispersion direction. The width of the profile is held fixed at a value which is determined globally for each plate. The maximum of the fitted profile curve is taken as extraction value.

`ext` Spectra of sources which appear extended on the direct plates are extracted in a similar manner as point-like sources, but the width of the spectral profile is *not* held fixed during the fitting procedure, and more pixels are used for the fit. Note that `ext`-sources are not necessarily galaxies. For example, if point-like sources are located in diffraction spikes of very bright objects or when two objects are located very close to each other, so that they overlap on the direct plates, they are classified as extended.

`bright` Spectra of objects close to saturation are extracted by *summing* the density values in each column perpendicular to the dispersion direction. This results in an extension of the dynamic range of the objective-prism plates by 2–3 magnitudes.

Each extracted, one-dimensional spectrum consists of 300 pixels. 20 pixels redward of the red edge are included, and 280 pixels blueward (towards *increasing* scan length `x`). Pixel values are densities D above diffuse plate background (bgr) in arbitrary units called counts. The relation between counts and photographic densities $D_{\rm photo}$ is

$$D_{\rm photo} = \frac{D\,[\rm counts] + D_{\rm bgr}\,[\rm counts]}{800}. \tag{2}$$

The HES spectra are stored as MIDAS frames, in which each row represents one spectrum. Each frame is accompanied by a MIDAS table, in which additional information for each spectrum is stored row-wise, e.g. object position on the sky, magnitude, and values for spectral features (see Section 3.2).

The mean number of overlap-free spectra with $S/N > 5$ is ~ 10000 per plate, so that the data base of the HES (including the 54 plates not yet prepared for stellar work) consists of ~ 4 million usable digital spectra. The number of spectra present on the 329 plates used in this work (effective area $\sim 6400\,{\rm deg}^2$ is 3 437 630.

2.2.5 Wavelength Calibration

A global dispersion relation for all HES plates has been determined by Ikonomou (1995), using A-type stars. In HES spectra of these stars the Balmer lines at least up to H_{10} are resolved (see Fig. 4), so that a dispersion relation can be computed by comparing the x-positions (scan length in μ m) of these lines, determined by template matching, with the known wavelengths. The wavelength calibration zero point is specified by the astrometric transformation between direct plates and spectral plates. The relation between λ[Å] and $x\,[\mu\,\mathrm{m}]$ is given by the following expression:

$$x = -2953.588 + 8.218377 \cdot 10^{10} \cdot \frac{1}{\lambda^2} + 7.675455 \cdot 10^{16} \cdot \frac{1}{\lambda^4} \tag{3}$$

The wavelength calibration is accurate to $\pm 10\mu$ m. This corresponds to ± 13 Å at Hγ and ± 5 Å at $\lambda = 3500$ Å.

2.2.6 Estimation of the Amplitude of Pixel-Wise Noise

The amplitude of pixel-wise noise as a function of density D is determined plate-wise using A-type stars (see also Christlieb 1995). A straight line fit is done to the spectral region between Hβ and Hγ (see Fig. 4). The scatter around the fit is purely due to noise, since in A-type stars the spectral region under consideration is free of absorption lines at the spectral resolution of the HES.

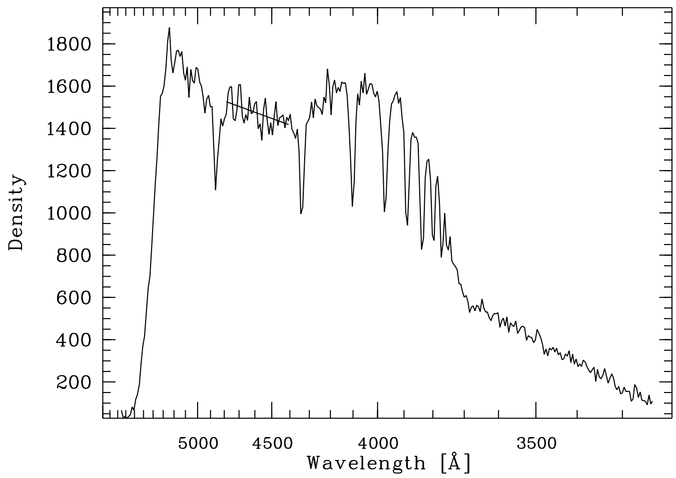

Figure 4: Measurement of noise in absorption line free spectral region of A-type stars. Note that we plot HES objective prism spectra such that wavelengths are *decreasing* towards the right, because the scan length x is *increasing* in this direction. The sharp drop of the spectra at the red end is due to the IIIa-J emulsion sensitivity cutoff at ~ 5400 Å.

We measure the amplitude of the scatter for all A-type stars present on each HES plate (typically ~ 50 per plate), and compute the mean density in the fit region. This yields data points (D, noise), to which a 2nd order polynomial is fitted, i.e.

$$\mathrm{noise} = a_0 + a_1 \cdot D + a_2 \cdot D^2. \tag{4}$$

We use a robust fit algorithm minimizing the sum of absolute deviations, and the following set of boundary conditions:

$a_0 > 0$. The noise at $D = 0$ is the noise of the plate background, which is always > 0.

$a_2 \geq 0$. Since $D \geq 0$, and the noise increases monotonically with D, a_2 must be positive (or zero).

$a_1 \geq 0$. From the previous boundary condition follows that the polynomial has a minimum at $D_{\mathrm{min}} = -a_1/(2a_2)$. Using again the argument that the noise increases monotonically with D, it follows that $D_{\mathrm{min}} \leq 0$. Since $a_2 \geq 0$, a_1 must be ≥ 0.

An example for a such a fit is shown in Fig. 5.

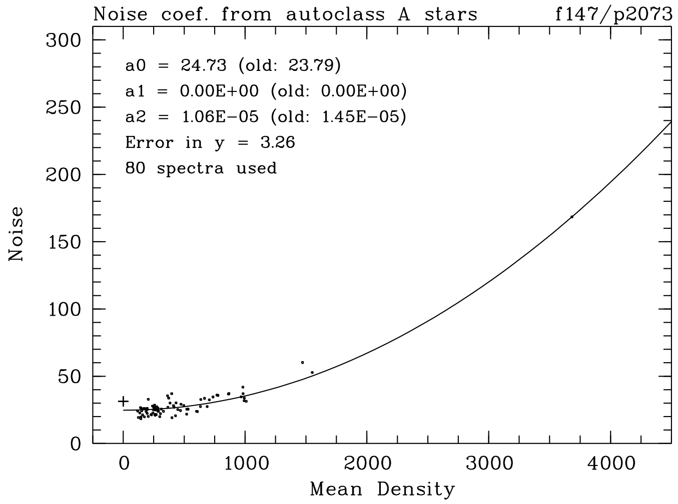

Figure 5: Fitting of a 2nd order polynomial to data points (D, noise), for estimation of pixel-wise noise as a function of density. The plate background noise is marked with '+'. It is always higher than expected from the noise parabola fit, since the optimal spectrum extraction algorithms result in a bias towards lower noise.

For simulations of spectra (see Sect. 3.1) it is very important to know the *form* of the distribution of noise. We investigated this by using 50 spectra of A-type stars from 8 plates with high sky background ($D > 1\,500$), and 60 spectra from 7 plates with low background ($D < 700$) – see Tab. 1. These spectra were chosen by hand from the sample of automatically selected A-type stars, in order to ensure that misclassified spectra, and spectra for which the fit of the continuum between Hβ and Hγ by a straight line is not fully adequate, do not confuse the results. 5 of the original set of 115 spectra have been excluded in the manual selection process.

The deviations from the continuum fits were collected for each spectrum, shifted to a median of 0, and divided by the average of the absolute values of the upper and lower 50 % quartile, so that a comparison of the noise distributions measured in different spectra (with different noise amplitude) is possible. The result is that the distribution of pixel-wise noise is almost perfectly Gaussian, independently of plate background (see Fig. 6).

An approximate relation between average S/N in the B_J band and B_J magnitudes has been derived by using 507 spectra from many different plates. Overlapping spectra undetected by the automatic overlap recognition algorithm, and spectra from plates with unusually high sky background have been excluded from the fit by iterative 3σ clipping. The resulting relation is

$$B_J = 16.8 - 0.086 \cdot \overline{\left(\frac{S}{N}\right)_{B_J}}. \tag{5}$$

Plate	Field	`bgr`	`bgrnoise`	$N_{\rm spc}$
10882	540	1645	31.5	6
11535	823	1728	34.7	9
11552	826	1925	39.4	6
11576	827	1790	36.4	8
11578	825	1541	31.9	2
11579	759	1836	38.3	2
11588	605	1608	32.6	6
12024	195	1555	33.8	11
10144	155	559	18.1	11
10846	539	644	18.3	6
10847	757	688	18.3	9
10849	422	583	17.5	12
11621	154	673	19.7	7
11622	300	665	19.4	7
11623	199	494	17.5	8

Table 1: HES plates used for the investigation of the distribution of pixel-wise noise.

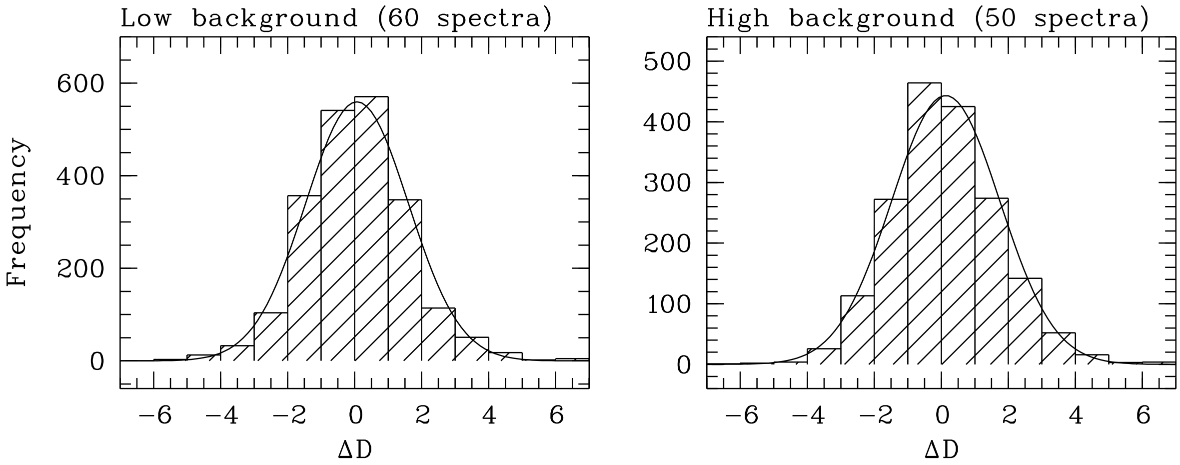

Figure 6: Distribution of pixel-wise noise in absorption line free regions of 60 spectra of A-type stars from 7 plates with low sky background (left panel), and 50 spectra from 8 plates with high background (right panel).

It has a 1σ error of 0.46^m. The large scatter is due to the fact that B_J for a given S/N can vary dramatically with plate background, and seeing. However, the purpose of the above relation is just to have a rough orientation of the B_J magnitude corresponding to a given S/N.

2.3 Candidate Selection

In the HES, the selection of candidates for any type of objects is carried out in two steps. In the first step, candidates are selected automatically by applying e.g. colour criteria, by automatic classification (see Sect. 3; see also Christlieb et al. 1997, 1998*a*,*b*), or other techniques. These candidates have to be

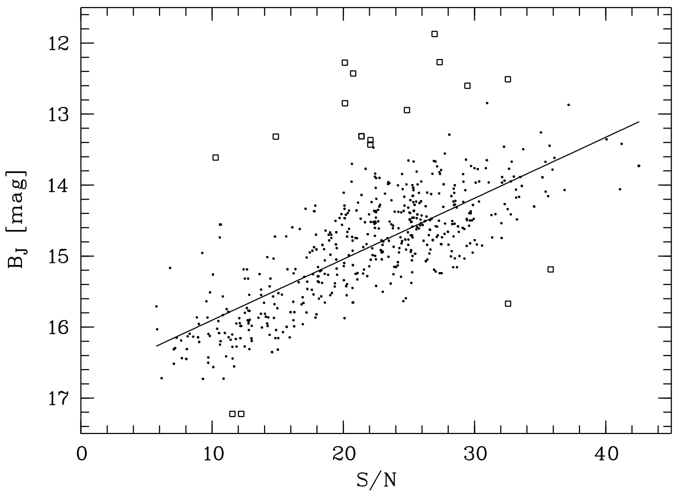

Figure 7: B_J as a function of average pixel-wise S/N in the B_J band. Open squares denote spectra that have been excluded from the fit by iterative 3σ clipping.

inspected manually to identify spectra which have been erroneously selected due to presence of plate artifacts (e.g. scratches, dust, emulsion flaws, ghosts, etc.). Moreover, a few overlaps escape detection by the automatic procedure, so that these spectra have to be rejected by hand. For these purposes, the program `screenspc` has been written. It displays the candidate spectra one by one, and the relevant regions of the direct plate and spectral plate scans upon request (see Fig. 8). With the help of the information being displayed on the screen, it is also possible to reproduce why the given spectrum has been automatically selected.

Acknowledgements

I would like to emphasize that the HES data base is the result of a team effort. The list of team members is too long to mention all of them here (however, see the footnote on p. 6). Special credits go to D. Reimers, the P.I. of the HES and L. Wisotzki for making the survey as successful as it is, the conception of the HES data reduction, and development of methods and algorithms for the exploitation of the data base.

References

Christlieb, N. (1995), Automatische Klassifikation von digitalisierten Objektivprismen-Platten, Diplomarbeit, Universität Hamburg.

Christlieb, N., Graßhoff, G., Nelke, A. & Wisotzki, L. (1997), Automatic classification of digitized objective prism spectra, *in* E. Kontizas, M. Kontizas, D. Morgan & G. Vettolani, eds, 'Wide-Field Spectroscopy', Kluwer, Dordrecht, pp. 109–113.

Christlieb, N., Graßhoff, G., Nelke, A., Schlemminger, A. & Wisotzki, L. (1998*a*), Automatic spectral classification, *in* I. Balderjahn, R. Mathar & M. Schader, eds, 'Classification, Data Analysis, and Data Highways', Springer, Berlin, pp. 16–23.

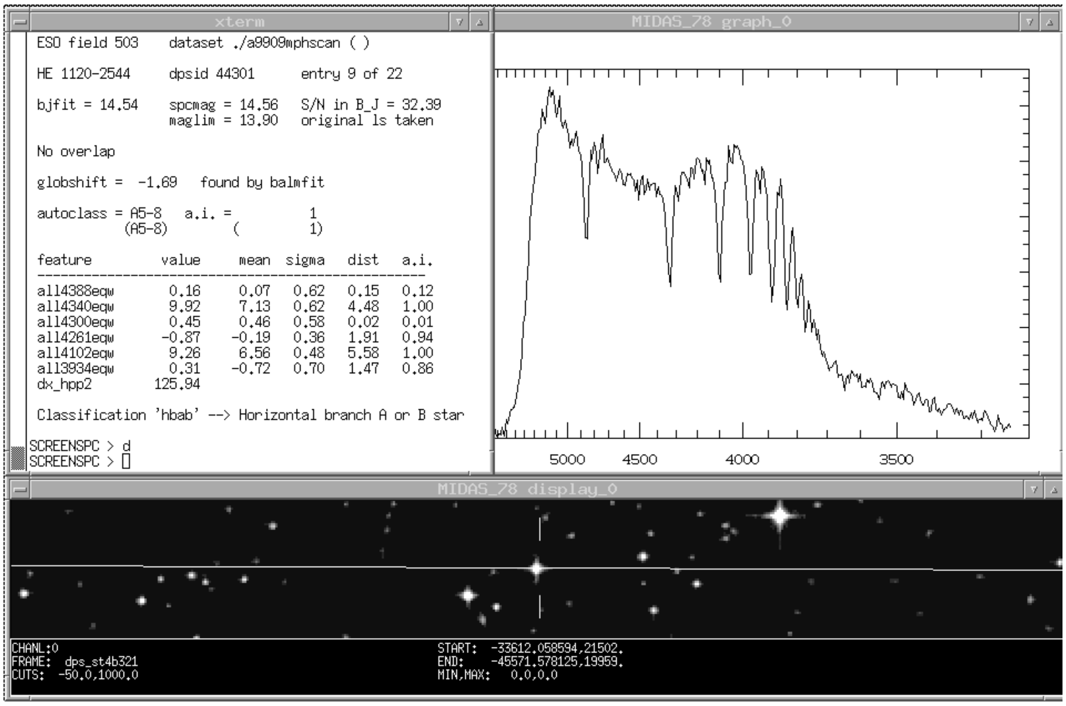

Figure 8: Screen-dump of the program `screenspc`, adapted for automatically classified spectra, in action. The information displayed in the upper left window allows to reproduce why the given spectrum has been assigned to its class (in this case class "A5-8") by automatic classification. In the upper right window the spectrum is shown. In the lower window the relevant regions of the direct plate and spectral plate scans can be displayed upon request in order to identify overlapping spectra and plate artifacts. The white line in the direct plate image corresponds to the dispersion direction of the spectral plate. In the middle, marked with vertical lines, the object under investigation can be seen. Direct and spectral image are displayed on the same scale.

Christlieb, N., Graßhoff, G., Nelke, A., Schlemminger, A. & Wisotzki, L. (1998*b*), Linné, a software system for automatic classification, *in* R. Albrecht, R. Hook & H. Bushouse, eds, 'Astronomical Data Analysis and Software Systems VII', Vol. 145 of *ASP Conf. Ser.*, pp. 457–460.

Deutsch, E. W. (1994), 'Positions and Proper Motions of Dwarf Carbon Stars', *PASP* **106**, 1134–1137.

Dickey, J. & Lockman, F. (1990), 'H I in the Galaxy', *ARA&A* **28**, 215–261.

Hagen, H.-J. (1987), Automatisches Suchen von Quasar-Kandidaten auf Objektivprismen-Platten, PhD thesis, Universität Hamburg.

Hewett, P. C., Foltz, C. B. & Chaffee, F. H. (1995), 'The Large, Bright QSO Survey. VI. Quasar Catalog and survey parameters', *AJ* **109**, 1498–1521.

Ikonomou, M. (1995), Untersuchung der Umgebung von Seyfert-Galaxien, Diplomarbeit, Universität Hamburg, Hamburg.

Köhler, T. (1991), Interne Helligkeitskalibration von Objektivprismenplatten, Diplomarbeit, Universität Hamburg, Hamburg.

Vanelle, C. (1996), Photometrische Eichung von Schmidtplatten, Diplomarbeit, Universität Hamburg, 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.

Wisotzki, L., Köhler, T., Groote, D. & Reimers, D. (1996), 'The Hamburg/ESO survey for bright QSOs I. Survey design and candidate selection procedure', *A&AS* **115**, 227–233.