# Chapter 3

# Preprocessing

Because the echelle spectrograph produces two-dimensional images on the CCD, many steps are required to produce the one-dimensional spectra in which the absorption lines can be seen and measured. To prepare the two-dimensional frames for this so-called data reduction, the digitized voltage signals—which are recorded by the CCD as data numbers (DN) or analog-to-digital units (ADU)—must be converted back into the number of photoelectrons or photons that created the signal in the first place.

## 3.1 ESO-MIDAS

The preprocessing and reduction of the data was performed almost entirely using the European Southern Observatory-Munich Image Data Analysis System. ESO-MIDAS, or MIDAS for short, is a software package containing general tools for image processing and data reduction (ESO-MIDAS User Guide, Vol. A).

The two-dimensional raw CCD images, also called frames in MIDAS, can be viewed in a display window, while plots of the pixels values versus the $x$- or $y$-coordinates can be generated in a graphics window. The preprocessing here made use primarily of the basic MIDAS commands for performing mathematical operations on the frames, such as the pixel-by-pixel subtraction of one frame from another. Later, during the data reduction, more use was made of higher-level command procedures with different options (see chapter 4).

## 3.2 Bias Subtraction

In order to keep readout noise from driving the analog-to-digital input below zero, the readout electronics of a CCD will always add a positive, constant offset to the signal from each pixel before it is digitized. This additive constant must therefore be determined and subtracted from all of the exposures before any further calibration steps can be done. Since the bias is present even when no photons are detected, the

bias level can be determined by taking the shortest possible (nominally zero-second) exposures with the shutter closed.

For high-quality CCDs, the bias is constant in time, remaining the same from one exposure to another. It also shows no spatial dependence; i.e., the variation in the bias value from pixel to pixel is significantly smaller than the readout noise in each pixel (Wisotzki 1994). In that case, one can combine all of the bias frames and obtain a master bias frame or merely a constant value that is then subtracted from all of the other CCD exposures. According to simple error propagation formulas, averaging several bias frames has the advantage of reducing the estimated error by a factor of $\sqrt{n}$, where $n$ is the number of frames used.

Unfortunately, the bias frames taken for HE 0940 show levels that vary not only from night to night but also over the surface of the CCD as well (see Figure 3.1). To compensate for the fact that the bias changed over time, a bias level was determined individually for each frame by averaging all of the pixel values in a stable part of the overscan region on the CCD. Between 191 and 210 ADU were thus subtracted from all of the frames in this first part of the bias correction.

After the information from the overscan region had been used for the first part of the bias correction, the edges were trimmed from all frames to avoid problems that could result from including areas with pathological behavior in the smoothing steps that followed. The START and STEP descriptors were then reset to (1.0, 1.0) so there would be no confusion between the so-called world coordinates and pixel coordinates.

Despite being scaled to about the same level, the spatial variations in the bias across each frame still remained. The strategy was to reduce them by subtracting a two-dimensional bias frame that had been strongly smoothed beforehand to minimize the effects of the readout noise.

The bias calibration frame for the exposures taken on the first night (Table 2.2) was created as follows. The top five curves in Figure 3.1 each represent the average of 5 individual bias frames. Since their pixel values and spatial structures overlapped well after subtraction of the overscan values, they were all averaged together and median filtered to remove most of the cosmics. The filtered frame was smoothed by fitting a polynomial of third degree in both the $x$ and $y$ positions on the CCD. This frame was then subtracted from frames casp0034.bdf and below.

The bias calibration frame for the second night's exposures corresponds to the bottom curve in Figure 3.1. It was created by averaging the remaining 5 bias frames, applying the MIDAS command `FILTER/COSMIC` (which was more effective than the median filter in removing a large cosmic covering four pixels), and again doing a two-dimensional, third-degree polynomial fit. This fitted frame was subtracted from frames casp0067.bdf and above.

The two-dimensional bias subtraction was reasonably successful in reducing most of the structure produced by a non-constant bias. The left panel of Figure 3.2 shows

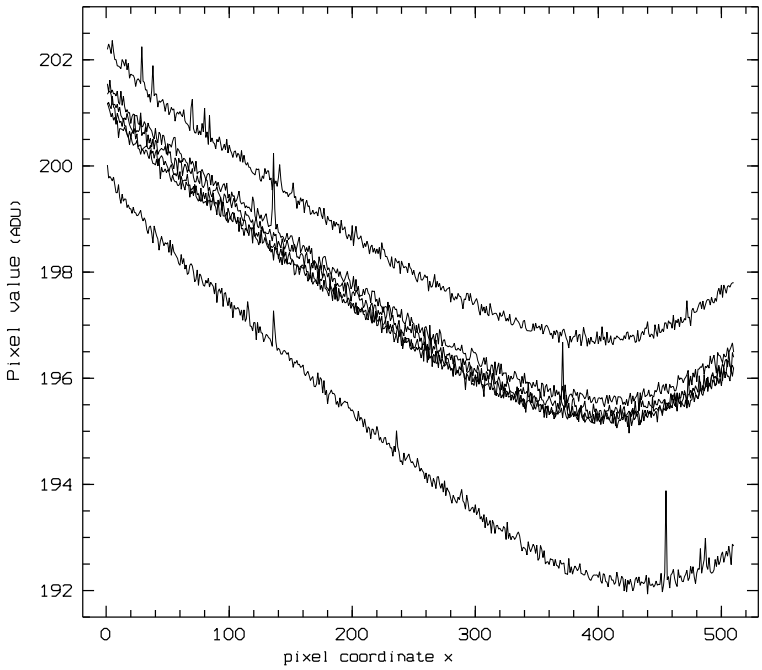


**Figure 3.1:** *Two-dimensional structure of bias frames as seen in the average column of each frame, shown before subtraction of the overscan value.*

the average column in each of four long dark frames before the bias correction (the cosmics have been filtered out for presentation purposes, to allow the structure to be seen). The right panel shows three of the average columns after the two-step bias correction. Note that casp0045.bdf is no longer present; it was discarded because its dark-current level was notably higher and its structure slightly different from that of the others. Since the dark frames were taken during the day and the detector—which is not completely light-tight—was left unattended, it is quite possible that scattered light degraded that exposure.

## 3.3 Dark-Current Correction

Independent of bias, the random thermal motions of atoms within the silicon lattice will produce enough energy to transfer some electrons from the valence band to the conduction band of the CCD, thus generating a non-zero signal even in the absence of light. Cooling the CCD or operating it in inversion mode (McLean 1997) greatly reduces but does not elimate this dark current, which adds charge linearly with exposure time and can be determined by taking long exposures with the shutter closed. The correction is then made to each non-bias frame by multiplying the dark current in ADU per second by the frame's exposure time in seconds, and subtracting the result from that frame.

As with any long exposures, the so-called dark frames show discrete bright spots

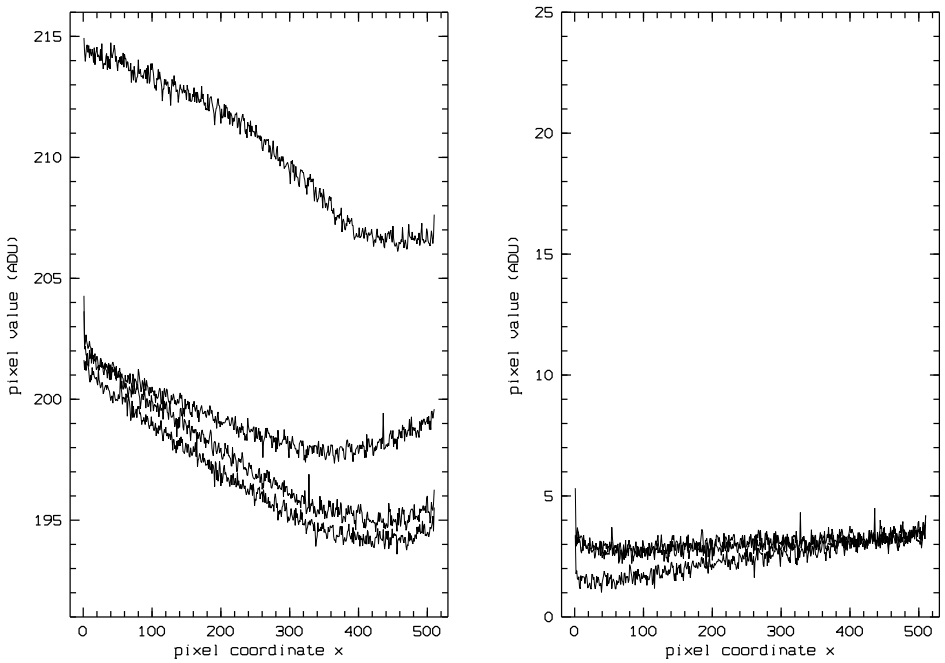


**Figure 3.2:** *Average filtered column of each dark frame. LEFT PANEL: before the two-step bias correction. RIGHT PANEL: after the bias correction (without casp0045).*

or small streaks caused by cosmic radiation—subatomic particles that deposit large amounts of energy when they are stopped in the CCD material. This happens up to a handful of times per $cm^2$ per minute, depending on the altitude. Given the low probability of two particles striking the same spot on the CCD in two different exposures, such cosmics can be removed by combining two or more long dark frames. But since casp0045.bdf had been discarded, only the long dark frame casp0034.bdf was available to calibrate the frames taken on the first night. Its counts differed from the long darks taken on the second night, so combining frames was not an option in this case.

Rather than attempting to remove all of the cosmics in casp0034.bdf by changing the pixel values manually with MIDAS commands such as `MODIFY/AREA`, a different approach was taken. The dark current was determined by making a histogram of the number of pixels vs. count rate in each of the remaining three long dark exposures (Fig. 3.3). Pixels containing cosmic events had much higher pixel values and were much less numerous than the 99.7% of lower-valued pixels that contained only dark-current counts, so the entries with cosmics were scattered well off to the right of the main distribution and thus neglected. A Gaussian was fitted to the histogram of dark-current counts, and its center value yielded the average dark-current value for each frame. The estimated dark current was 1.9 ADU per hour on the first night. The corresponding results for the other two dark frames were averaged to yield 2.5 ADU per hour for the second night.

Although the dark current seemed negligible at about 0.0006 counts per second, the long exposure times needed for the quasar images meant that it accounted for 8% and 11% of the count rates in the object frames taken on the first and second night, respectively.

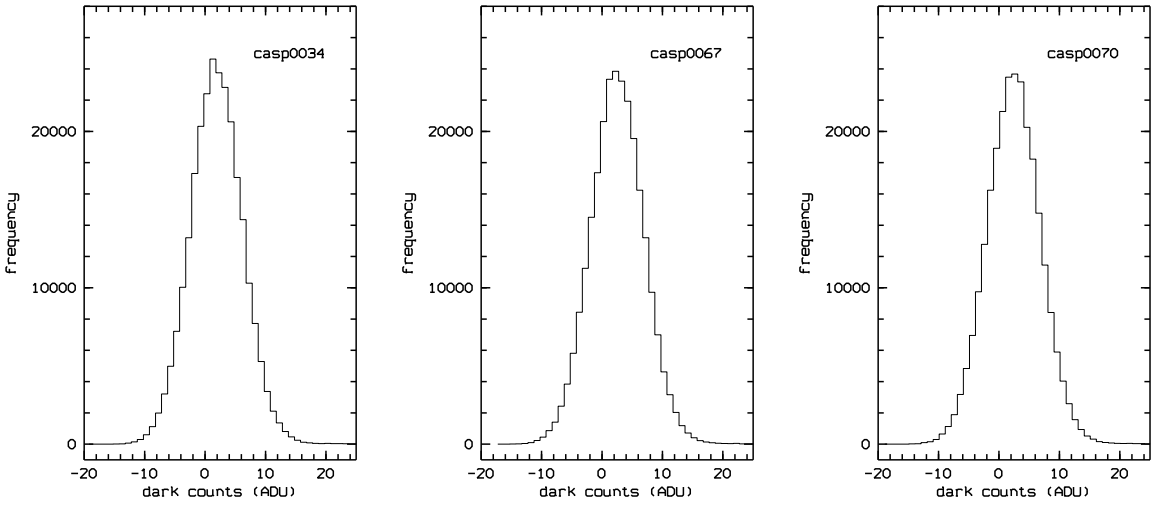


**Figure 3.3:** *Histograms of counts in long dark frames*

## 3.4 Conversion to Photoelectrons

When a photon strikes the CCD, its energy looses a certain number of electrons from the valence band into the conduction band, allowing them and the corresponding "holes" in the valence band to move through the material under the influence of an electrical field. The resulting voltage is read out and amplified by the detector electronics. The number of counts or ADU of the actual output signal $S$ is thus

$$S = \frac{N_e}{g} + d\,\Delta t + b \tag{3.1}$$

where $d$ is the dark current in units of ADU per second, $\Delta t$ is the integration time in seconds, $b$ is the bias in ADU, and $g$ is the system's photon transfer gain factor in electrons per ADU. Clearly, after the bias and dark current have been subtracted, the remaining signal is directly proportional to the number of photoelectrons $N_e = \eta N_\gamma$, where $N_\gamma$ is the number of photons that were absorbed in the time $\Delta t$. McLean (1997) describes in detail how the value of $g$ can be either calculated from the overall amplifier gain and the CCD capacitance, or measured experimentally. In this case the conversion factor was simply taken to be 2.1 electrons per ADU, following the information contained in the MIDAS descriptors of the data frames.

Poisson statistics apply to photon or electron counts but not to ADU, so this conversion was needed in order to calculate the photon noise; namely $\sqrt{N}$ for a signal of $N$ photoelectrons. On the other hand, the CCD electronics produce a certain amount of noise whenever they are read out, regardless of the size of the signal. In the case of CCD #32, this noise level was $\sim 11$ e$^-$ (Pasquini 1993). Assuming that the photon noise and the read-out noise $\sigma_r$ are independent and random, the expression for the combined noise in the raw frames becomes

$$\sigma = \sqrt{{\sigma_N}^2 + {\sigma_r}^2} = \sqrt{N + {\sigma_r}^2} \tag{3.2}$$

Additional errors arise as changes are made to the frames during preprocessing and data reduction. Since the bias is read out without a photon signal, the error associated with a constant bias level is due to the read-out noise and is thus equal to $\sigma_r$. The two-dimensional bias correction is also subject to the errors in the read-out noise, but error propagation yields an error of $\sigma_r/\sqrt{n}$ because $n$ frames were averaged to obtain each bias frame. In the end, the variances $\sigma^2$ for the raw frames, the bias correction, the dark-current subtraction and later the background subtraction all had to be added together and kept for later use in the optimal extraction of the spectrum (see section 4.5).