Chapter 4

Data Reduction

The purpose of the reduction is to convert the two-dimensional raw CCD images obtained at the telescope into a one-dimensional spectrum containing the measured intensity as a function of wavelength. Since this was one of the main tasks of this work—and certainly the most time-consuming—it will be described in detail in the following sections.

4.1 The Echelle Context in MIDAS

Most of the work described in this chapter was done within the MIDAS echelle context, which is an application package containing commands specially created for the reduction of echelle spectra. Information on the use of these commands was gleaned from chapter 7 and appendix D of the MIDAS manual, volume B.

An echelle session within MIDAS has a few additional features that should be mentioned briefly. During each reduction step that is performed within the echelle context, new information is written to temporary system tables which are automatically created in the current working directory. They are called order.tbl, back.tbl and (if any wavelength calibration steps were done) line.tbl. Since MIDAS refers to these particular tables in any subsequent reduction steps, it is usually necessary to preserve the information for use in a future MIDAS session by saving the tables under a session name.[1]

The echelle context uses a large number of variables, or MIDAS keywords, as session parameters to specify input frames, default command arguments, and reduction methods or options for each of the reduction steps. They are stored in the system tables and can be reviewed using the command `SHOW/ECHELLE`.

[1] For example, the command `SAVE/ECHELLE SESS` copies the temporary tables to the files SESSORDE.tbl, SESSback.tbl and SESSLINE.tbl. The information can later be made available via the command `INITIALIZE/ECHELLE SESS`, which copies the tables with this prefix back into the tables order.tbl, back.tbl and line.tbl.)

4.2 Order Definition

The first step in the data reduction is to define the positions of the spectral orders on the two-dimensional image. As described in chapter 7 of the MIDAS manual, this means solving the equation

$$y = f(x, m) \tag{4.1}$$

which relates the line number y to both the sample number x and the spectral order m on the raw image. The reduction software assumes that the images have been rotated and flipped so that they are oriented as follows: the spectral orders almost horizontal on the image, with a slight positive slope; the wavelength scale increasing from left to right; and the absolute order number increasing from top to bottom, so that the separation between spectral orders is smaller at the bottom that at the top.

The MIDAS command `DEFINE/HOUGH` was used to obtain the initial positions and slopes of the spectral orders on the CCD. In this algorithm, the center of each order is defined as the first moment of the pixel values above a certain threshold and is approximated by a polynomial

$$y \approx \sum_{j=0}^{J} \sum_{i=0}^{I} a_{ij}\, x^i m^j \tag{4.2}$$

The coefficients a_{ij} which define the order positions are stored in the table order.tbl. Points belonging to the background between the orders are also defined at this time and stored in the table back.tbl.

Because the quasar images of HE 0940 are so faint, the order definitions had to be done on the much brighter flat-field frames. This meant that when the spectrum was extracted from the wavelength-calibration (WLC), standard-star (STD) and quasar frames, the algorithm used the initial order positions obtained from the flat-fields taken just before or afterwards (see Table 4.1). All of the frames contain 43 echelle orders.

4.3 Background Subtraction

Even after subtraction of the bias and dark current, the pixels between the bright spectral orders still have non-zero values. This inter-order background, which is due to general scattered light from the instrumentation as well as diffuse light from adjacent orders, must be estimated and then subtracted from the entire frame.

The standard procedure uses the grid of background points identified and stored in back.tbl to estimate the inter-order background. This is done either by a polynomial fit of the selected points (method `POLY`) or by smoothing spline interpolation (method `SPLINE`), which divides the grid into discrete intervals and fits them with polynomials whose function values and derivatives are made to be continuous at the endpoints of the intervals where the adjacent polynomials meet. Both methods are implemented via the MIDAS command `BACKGROUND/ECHELLE`.

When applied to the quasar and flat-field frames, however, it turned out that the single polynomial in method `POLY` was not flexible enough to follow the background level in all cases, while method `SPLINE` produced artifacts and uneven patterns on the edges of the image. Only the command `BACKGROUND/SMOOTH`—which uses an iterative algorithm to heavily smooth the frame perpendicular to the orders and replace higher pixel values with the smoothed values—was able to produce background estimates that showed no undesirable features but had reasonable levels. Because read-out noise causes fluctuations even where there would otherwise be a constant level, the estimated background should not pass through the very lowest pixel values, but rather more or less through the middle of those fluctuations. Figure 4.1 shows an example of this.

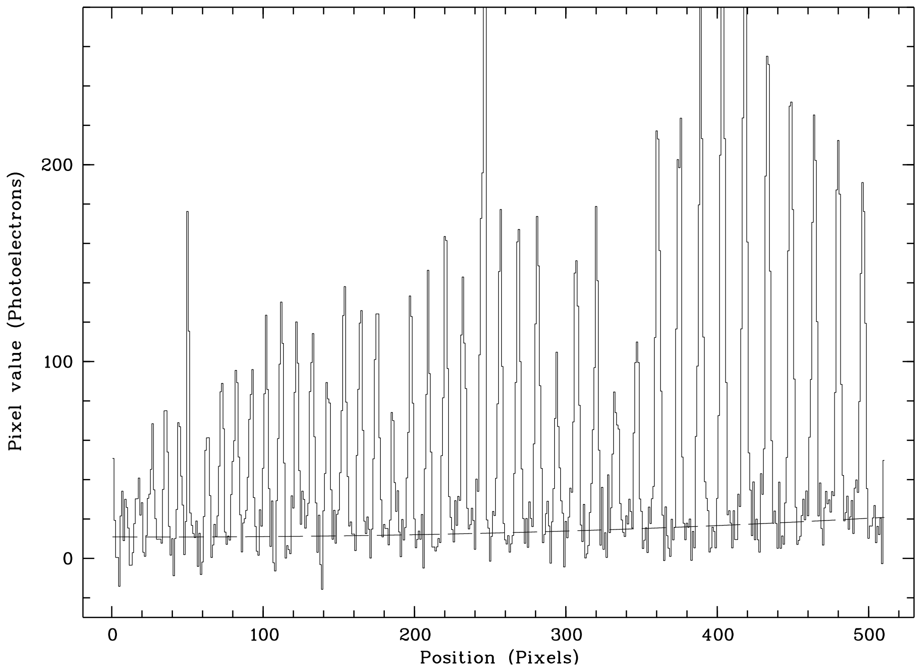

Figure 4.1: *Sample column in quasar frame casp0031 (solid line) and estimated inter-order background level (dashed line).*

The main disadvantage of `BACKGROUND/SMOOTH` was that it produced an extremely "lumpy" background for the wavelength-calibration frames; but since the identification of emission lines was not impaired, this was not judged to be critical. Thus the inter-order background in the WLC, standard-star and quasar frames was estimated via the smoothing method and then subtracted.

4.4 Flat-field Correction

The primary purpose of flat-field frames is to correct for the non-uniform sensitivity across the CCD. Flat-fields are created by exposing the CCD to a uniformly illuminated field, such as the one provided by the inside of the white telescope dome or a flat white screen uniformly illuminated by lamps; exposures of this kind are called dome flats.

Following McLean (1997), let I be the uniform illumination falling on a pixel in row i, column j whose quantum efficiency is η_{ij}. After the bias, dark current and inter-order background have all been subtracted, the observed signal in ADU from that pixel in the flat-field frame is

$$(X_{ij})_{FF} = \frac{1}{g}\ \eta_{ij} I. \tag{4.3}$$

The averaged, or sometimes merely strongly smoothed, flat-field has pixel values

$$\langle (X_{ij})_{FF} \rangle = \frac{1}{g}\ \eta_m I \tag{4.4}$$

where η_m is the mean quantum efficiency averaged over the entire image. When the original flat-field frame is divided by this average, the resulting "normalized" flat-field calibration frame

$$(X_{ij})_{cal} = \frac{(X_{ij})_{FF}}{\langle (X_{ij})_{FF} \rangle} = \frac{\eta_{ij}}{\eta_m} \tag{4.5}$$

has pixel values in the neighborhood of 1 which are essentially multiplicative factors representing the pixel sensitivity.

In a standard-star or science frame, the illumination I_{ij} is no longer constant and the observed signal from the pixel in row i, column j in a dark- and background-subtracted frame is

$$X_{ij} = \frac{1}{g}\ \eta_{ij} I_{ij} \tag{4.6}$$

The pixel-to-pixel variations represented by the factor η_{ij} can now be removed by dividing the frame in question by the normalized flat-field, yielding an image whose pixel values contain the relative intensity

$$\frac{X_{ij}}{(X_{ij})_{cal}} = \left(\frac{\eta_m}{g}\right)\ I_{ij} \tag{4.7}$$

which differs from the true illumination I_{ij} by only a constant factor. If necessary, the scale factor η_m/g can be determined by observing a source of known brightness and scaling the pixel values to match the known absolute flux.

Telescope flexure leads to changes in the CCD position of the infalling light over the course of the night, so in keeping with the usual practice flat-fields were taken immediately before and after each object exposure to ensure that the order positions would not differ much. Unfortunately, although the order centers obtained from the flat-fields matched those in the quasar frames well, the quasar orders themselves were wider because of slight shifts in position. Had the flat-fields been taken using a longer slit, the quasar and flat-field orders might have overlapped fully. But as it was, the division shown in equation (4.7) could not be carried out because it would have led to undefined regions in the QSO spectrum.

The flat-field correction therefore had to be neglected, and the flat-fields could be used only for order definition. Luckily, the CCD images contained no obvious features such as the fringes that sometimes appear in thinned CCDs due to interference from the night-sky emission lines.

4.5 Spectrum Extraction

Normally, the echelle orders on the CCD image are extracted by summing or averaging the pixel values in the y direction along a "numerical slit" one pixel wide, whose height is set by the user. The sum of the pixel values yields the flux as a function of x for each spectral order on the CCD. In practical terms, this means converting a two-dimensional raw image in pixel-pixel space to a two-dimensional "extracted" image in pixel-order space, i.e., whose x-axis is still the pixel coordinate but whose y-axis is now the relative order number. However, the final signal-to-noise ratio (S/N) is not always optimal. In particular, the many cosmics in long exposures must be removed manually.

In order to extract the spectrum in an optimal manner that maximizes the S/N, Sebastián López at the Hamburg Observatory has written a series of procedures that were used here instead of the standard MIDAS command `EXTRACT/ECHELLE`. In addition to a more flexible order definition, the main advantage of the Hamburg programs is that they take the variance of each pixel into account. The required input is therefore not only the usual frame containing the orders but also its variance frame, i.e., a frame whose pixel values equal the variance of the corresponding pixels in the frame that is to be extracted.

The variances for the frames processed here included not only the combined errors from photon noise and read-out noise, but also the estimated errors from the bias, dark-current and background corrections. Just before the extraction, cosmics were identified either as pixels for which the ratio of the input frame to a filtered version of the input frame exceeded a pre-defined threshold, or as pixels having more than the maximum count rate chosen for each frame by inspection. These outliers were given practically infinite variances.

For each echelle order, the extraction program used the positions provided by the MIDAS order definition (including any user-defined constant offset) as the initial values. At each x, a Levenberg-Marquardt algorithm fitted Gaussians to the order profile in the y direction along the slit; the pixels were weighted according to their variances. Once the Gaussian fits had yielded the order center in y at each x, a polynomial fit through the order centers was used to fix the position of the order as a whole. Then the spatial profiles were fitted once again, but this time the position and FWHM of the Gaussians were held fixed and only the amplitudes were allowed to vary. The area under the resulting curve at each x represented the extracted flux and was written to the output frame, while the error in the flux was written to a corresponding error frame. The flux errors proved extremely useful when trying to combine the three quasar spectra in a sensible way (see section 5.1).

The spectrum extraction achieved good results, as can be seen in Figure 4.2. Because of their long exposure times, the science frames in particular had been plagued by cosmics; but since cosmics had been tagged beforehand by means of variance frames, the Hamburg programs gave them correspondingly low weights in the Gaussian fits. Thus the optimally extracted flux values showed far fewer cosmics than those obtained with `EXTRACT/ECHELLE`.

Finally, all of the extracted frames were divided by their exposure times. This was in part because the MIDAS command used as the basis for the subsequent response correction assumed that the counts are normalized to an exposure time of one second.

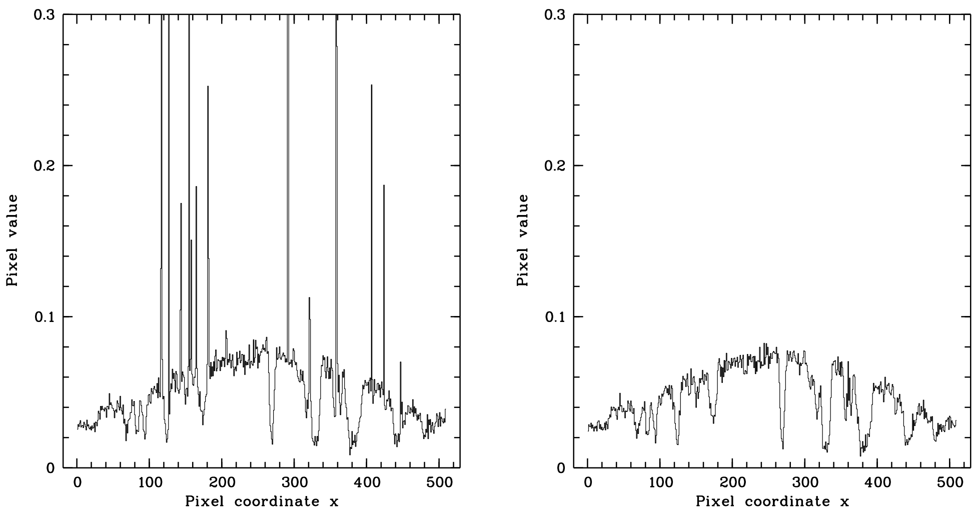

Figure 4.2: *Sample order (25) extracted from background-subtracted quasar frame casp0089.bdf. LEFT: result of standard MIDAS extraction. RIGHT: result of of Hamburg programs.*

4.6 Wavelength Calibration

The spectrum extraction provides flux values as a function of the sample position x on the CCD and the spectral order m. A wavelength calibration is still needed to translate the x position within each order into the actual wavelength λ; that is, the relation

$$\lambda = f(x, m) \tag{4.8}$$

must be solved together with equation (4.1) in order to obtain the desired transformation between (x, y) space and (λ, m) space. This is done with the help of wavelength calibration (WLC) frames, which in this case were exposures of a thorium-argon arc lamp.

The wavelength calibration in the MIDAS echelle context is done in two steps. The first step involves the extracted WLC frame, in which the thresholding algorithm in `SEARCH/ECHELLE` automatically detects the positions of emission lines. The user determines whether the center of each line should be estimated by fitting a Gaussian to the line profile (which was the method used in this case) or by finding its center of gravity. The line centers in x are then stored in the table line.tbl.

The second step in the wavelength calibration requires the interactive identification of two or more emission lines in the raw, or preferably background-subtracted, WLC frame. These lines must also have been successfuly detected by `SEARCH/ECHELLE` in the extracted frame. Since the user is asked to provide the laboratory wavelengths of the calibration lines as well as the absolute echelle order number of the first line, it is necessary to choose and identify these lines beforehand. This was done by plotting the intensities along an order in the two-dimensional frame (e.g., via `EXTRACT/TRACE`) and comparing them to the plots in the atlas of the thorium-argon spectrum (S. D'Odorico et al. 1987). Once certain groups of lines were recognized, it was easy to confirm that the CASPEC exposures at hand cover the spectral orders from 108 to 150.

The first wavelength calibration was performed interactively on frame casp0030.bdf using the `IDENTIFY/ECHELLE` method `PAIR`, which exploited the fact that the overlap in the wavelength regions covered by the spectral orders caused some emission lines at the red end of one order to appear again at the blue end of the next order. The algorithm uses the four locations of two such lines, chosen by cursor, to derive a set of global dispersion coefficients in an initial, third-degree polynomial approximation of the relation in equation (4.8). An iterative loop improves on the global relation while it identifies more lines by computing the estimated wavelengths and comparing them to the catalog wavelengths. When no more new lines can be identified, the algorithm goes on to compute dispersion coefficients and residuals for each order individually and writes them to the table line.tbl. The calculated wavelength step, i.e., the number of angstroms per pixel, is written to the line.tbl descriptor PIXEL as well as to the echelle session keyword SAMPLE, for use in the rebinning step (see below).

Since all of the WLC exposures were taken with the same instrumental set-up, the dispersion coefficients derived from casp0030.bdf could then be used as initial values in the other wavelength calibrations that followed. This was done using the `IDENTIFY/ECHELLE` method `GUESS`, which eliminated the need for further user interaction. As in the first WLC frame, the wavelength step was determined to be 0.1267 Å per pixel in all of these cases.

The wavelength calibration on each of the four WLC frames listed in table 4.1 yielded acceptable results. `SEARCH/ECHELLE` located between 647 and 713 emission lines in each of these frames, and the percentage of identifications among the brightest half of the lines used by `IDENTIFY/ECHELLE` was between 92% and 94% in every case. The differences between the catalog wavelengths and those obtained from the fit of the dispersion relation for the lines selected by `IDENTIFY/ECHELLE` were generally less than 0.02 Å, or less than 1/6 of a pixel. The residuals in each order of the first WLC frame are shown in figure 4.3, as are the histograms of these values.

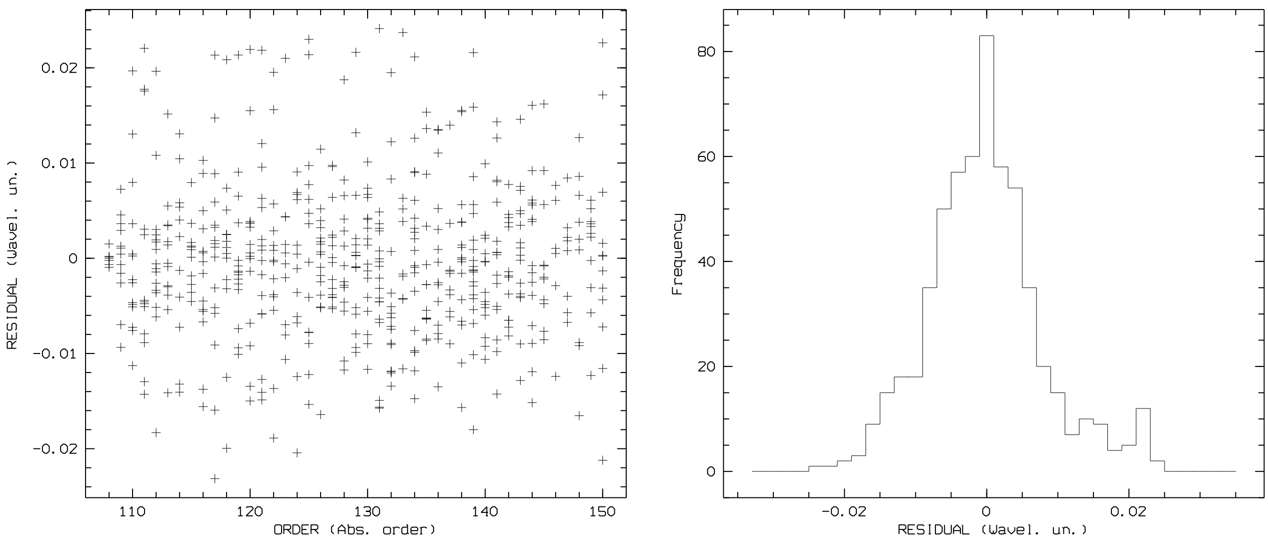

Figure 4.3: *Residuals from the wavelength calibration for casp0031 and histogram of residual values.*

4.7 Rebinning in Wavelengths

Once a wavelength calibration has been done, the stored dispersion coefficients and wavelength step can be used to convert any extracted frame into a so-called rebinned frame. The command `REBIN/ECHELLE` takes an extracted frame in pixel-order space and samples it in uniform wavelength steps, thus converting the pixel coordinates in each order of the input frame into wavelengths. The result is a rebinned frame in wavelength-order space, whose y-axis is still the relative order number but whose x-axis is now in wavelength units. The starting wavelength and the number of pixels in each order are written to the newly created descriptors WSTART and NPTOT in the output frame.

The two-dimensional rebinned MIDAS frame is, however, not the final result. Each spectral order that occupies a row in a rebinned frame has a different starting wavelength, yet shares the same pixel coordinates in x as all of the other orders. So the pixel coordinate x—though it is now in wavelength units—represents not the absolute wavelength but rather the relative wavelength, i.e., the distance in angstroms from the starting wavelength of each order. This is why a plot of one of the rows in a rebinned frame will show the flux values of that order on the y-axis, but not the absolute wavelength scale on the x-axis.

4.8 Order Merging (I)

The spectrum with the absolute wavelengths is obtained via the MIDAS command `MERGE/ECHELLE`, which takes a rebinned frame as input and produces a one-dimensional frame with the wavelengths along the x-axis and the measured fluxes contained in the pixel values. The algorithm first identifies the regions in which two orders overlap, then takes the weighted average of the pixels in those regions. Since no variance or error frames are used in the standard MIDAS echelle reduction, the weights are determined under the assumption that the closer a pixel is to the edge of an echelle order, the less reliable its value: the normalized weights are a linear function of position within the overlap, so that the pixel weights in the bluer order decrease with wavelength, in even steps from 1 to 0, while the pixel weights in the redder order increase steadily from 0 to 1. In the resulting complete spectrum, flux values in the overlapping regions covered by two orders are replaced by the weighted averages.

It was at this point that the complete arc-lamp spectra were obtained from the rebinned WLC frames. For the WLC exposures, this was the final step in the reduction process needed to produce usable one-dimensional spectra for later use (see section 6.1). For the QSO frames, further corrections required the use of the rebinned standard-star frames (see sections 4.9 and 4.10). The subsequent order merging for the QSo frames was also more complicated (sections 4.10 and 5.2).

4.9 Response Correction

Although flat-fielding corrects many of the variations in the pixel-to-pixel sensitivity of the CCD, two further factors affect the measured flux. The first is the echelle blaze effect, which results from variations in the efficiency of the spectrograph as a function of position along the echelle order; this is because the efficiency of the echelle grating is generally higher in the center of each order and decreases on edges. The second effect is the chromatic response of the instrument, which means simply that the detector sensitivity also varies as a function of wavelength. One reason is that photons of different wavelengths are absorbed at different depths in the CCD and the material is not homogeneous (McLean 1997). Because the color temperature of the artificial lamps used for dome flats differs greatly from that of the night sky, flat-fields of this type are not very effective in correcting for the blaze function and the instrument response, so the flux calibration must be done as an additional step.

Since the small slit length used to obtain the present exposures leads to a loss of light, there was no sense in attempting an absolute flux calibration. But the response correction was done anyway, to lessen the effects of the blaze function. This "flattened out" the orders, thus reducing the differences between the overlapping parts of adjacent orders and hopefully minimizing the size and occurrence of "bumps" at these overlaps when the spectrum was merged. The idea was that it would be easier to estimate and fit the continuum if the spectrum was flatter to begin with.

The method recommended in the MIDAS manual, which attempts to correct for both the blaze effect and the chromatic response at once, requires the use of a standard star whose known flux values are stored in a MIDAS table. Once the standard star has been observed with the same instrumental configuration and reduced in exactly the same way as the object, the correction factors can be calculated by comparing the relative flux values obtained from the standard-star frame with the absolute flux values listed in the table. Specifically, the command `RESPONSE/ECHELLE` begins by taking a standard-star frame as input. It subtracts the background, optionally performs the flat-field correction, extracts the spectrum, divides the extracted frame by the exposure time in seconds, and does the rebinning. Then a median filter is applied (in which pixel values that differ too much from the median of the surrounding pixels are replaced by that median value), followed by a smoothing filter (in this case similar to the median filter, but using the average rather than the median).

The algorithm continues by converting the values in the flux table into a frame, then dividing this frame by the rebinned and smoothed standard-star frame to produce the so-called response frame. The pixel values in the response frame represent the known flux divided by the observed flux; they are essentially the factors by which each pixel in an object frame should be scaled to yield the corrected flux.

As noted in the MIDAS manual, one source of possible inaccuracies is the difference in sampling step between flux table and standard-star frame. In this case the available flux table for μ Columbae was sampled at intervals of 9 Å, so its flux values had to be interpolated to the rebinned standard-star frame's smaller sampling step of 0.1267 Å before the frame division could be carried out.

A much more critical factor was the presence of absorption lines in the standard-star frame which are not in the table. These regions would have led to an incorrect response and had to be edited manually. This, and the fact that the standard-star extraction was done with the Hamburg procedures rather than with the usual MIDAS commands, required several modifications to the response correction described above.

First, the standard-star frame casp0078.bdf was background-corrected, extracted according to the Hamburg procedures, divided by the exposure time, and rebinned. The individual order spectra were obtained as 43 separate frames via `MERGE/ECHELLE NOAPPEND`. These order frames were converted into tables, where the regions without any absorption lines could be selected interactively and fit using fifth-degree polynomials. (The polynomials produced smoother curves and a closer fit than the median filter and smoothing normally done in `RESPONSE/ECHELLE`.) These order fits were free of both cosmics and absorption lines and were therefore suitable for calculating the response correction; but since the response frame is normally a rebinned frame, all 43 fits were copied into a two-dimensional frame with the same format and descriptors as a rebinned frame. The procedure `RESPONSE/ECHELLE` was modified so that it would use this optimally extracted, rebinned frame with the fitted orders rather than using the standard MIDAS procedures to reduce a raw standard-star frame from scratch. Thus it created the frame with the absolute fluxes as usual, but then divided that frame by the fitted rebinned frame to yield a response frame much less corrupted by extraneous absorption.

Figure 4.4 shows a sample standard-star order, along with the fifth-degree polynomial fit used in the response correction. Figure 4.5 shows the corresponding rows in the response frame, compared to what the response would have been if the absorption regions had not been replaced by the polynomial fits.

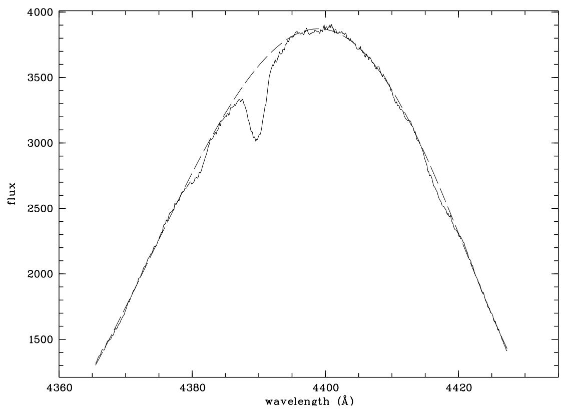

Figure 4.4: *Sample standard-star order 22 (solid line) and fifth-degree polynomial fit (dashed line).*

4.10 Reduction of the QSO Frames

After the wavelength calibrations had been done and the response frame created, the quasar frames could finally be reduced in one go.

First, the initial order definitions for each quasar frame were chosen from one of the appropriate flat-fields. As mentioned earlier, the flat-field and wavelength-calibration

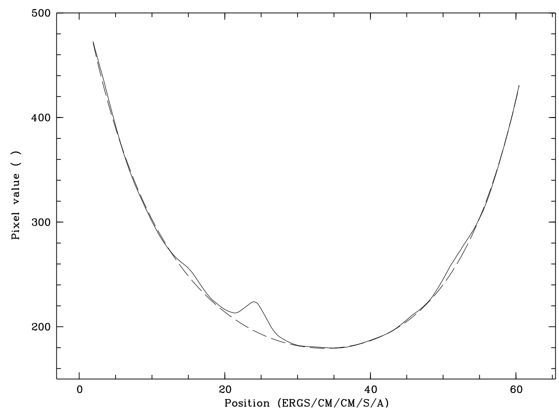

Figure 4.5: *Sample response function for order 22, uncorrected (solid line) and corrected for absorption (dashed line).*

frames used to make corrections were taken just before or after each quasar frame, to avoid introducing additional errors as a result of the positional shifts that take place over the course of the night. Table 4.1 summarizes the calibration frames used for each object exposure.

	QSO frame casp0023	QSO frame casp0031	STD frame casp0078	QSO frame casp0089
BIAS frames	casp0004–casp0007 casp0035–casp0039	casp0004–casp0007 casp0035–casp0039	casp0069	casp0069
DARK frames	casp0034	casp0034	casp0067 casp0070	casp0067 casp0070
FLAT (order def. only)	casp0025	casp0029	casp0077	casp0087
WLC	casp0024	casp0030	casp0076	casp0088
response	casp0078	casp0078	n.a.	casp0078

Table 4.1: *Object and Calibration Frames*

After the order definitions had been chosen, the inter-order background on all three quasar frames was estimated using `BACKGROUND/SMOOTH` and then subtracted. At this point the background-corrected frames would normally have been divided by the normalized flat-field, but this was not done for the reasons explained above.

A variance frame was now created for each of the quasar frames, and any cosmics that had been identified were tagged with practically infinite variances. The programs written in Hamburg made use of these variance frames to optimally extract the spectral orders from each of the background-corrected quasar frames. The program output was not only a frame in pixel-order space containing the extracted fluxes, but an equivalent frame containing the corresponding errors for each flux value. In all of the steps following the optimal extraction, the error frames were processed in parallel with the data frames, so that these error estimates were available at each step.

The extracted spectrum and the extracted errors obtained from the quasar frame casp0031.bdf were the first to be rebinned. It was first necessary to initialize the session with the appropriate wavelength calibration—and thus ensure that the correct dispersion coefficients were available in line.tbl—before executing the command `REBIN/ECHELLE`. The other two quasar spectra and their errors were also rebinned using the dispersion coefficients from the WLC session appropriate to each of them, but in those two cases the echelle keyword `SAMPLE` was used to define the first rebinned image as a reference frame. This guaranteed that all three spectra would have the same starting wavelength and wavelength step.

The response correction was applied to the three rebinned quasar frames (as well as their rebinned error frames) by simply multiplying each of them by the response frame. The effects of this correction on the fluxes can be seen by the example in Figure 4.6.

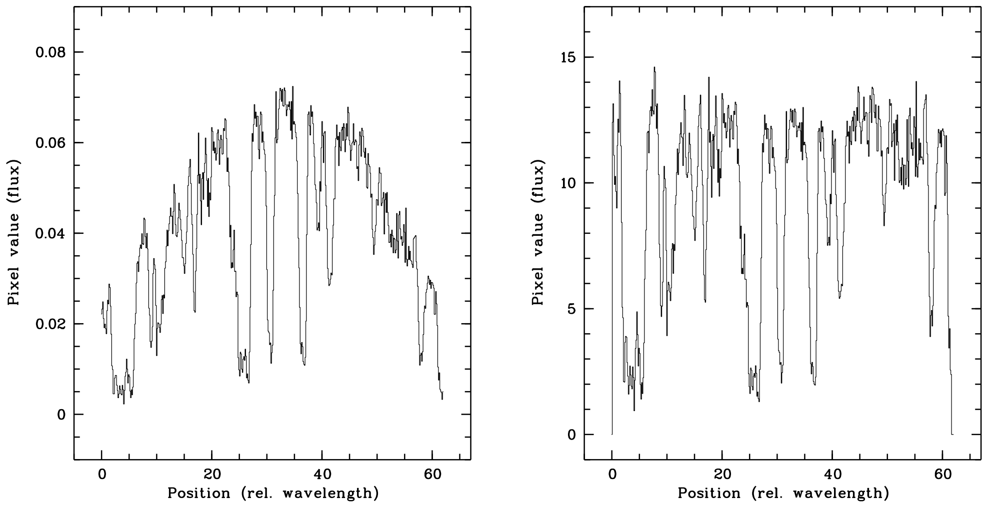

Figure 4.6: *Sample echelle order (22) before and after response correction.*

The final task in the normal MIDAS echelle reduction is to produce one-dimensional spectra from the response-corrected, rebinned frames. For reasons explained below, this was done in several steps. The first of these steps was to apply `MERGE/ECHELLE` to the data and error frames, but this time with the option NOAPPEND, so that nothing more was done than to take each of the orders from a rebinned frame, convert the relative wavelengths along the x-axis to absolute wavelengths, and store each order in a separate frame.

This concluded the main part of the data reduction, which produced three one-dimensional spectra for each of the 43 echelle orders from the three quasar exposures listed in Table 2.2.