Chapter 5

The Spectrum Toward HE 0940

Data reduction was defined as the production of one-dimensional spectra from two-dimensional CCD images. However, a number of technical steps were still necessary in order to achieve a spectrum suitable for fitting absorption lines. A single, normalized spectrum was needed, i.e., a spectrum in which the different quasar exposures had been combined, the orders merged, and the fluxes at each wavelength divided by the estimated continuum level there.

5.1 Combination of QSO Orders

At this point the spectra from each of the three quasar exposures were combined into one spectrum per order. This not only reduced the errors and increased the signal-to-noise-ratio, but also further minimized any traces of cosmics that might have survived the data reduction. Since the error frames produced during the extraction were also available, it made sense to take advantage of them and use the errors at each wavelength to determine the weighted average for each order.

If each of three data points x_i has been measured with the precision σ_i, then the most probable value is the weighted average of the data points, where the weight w_i given to each point is inversely proportional to the variance of that data point (Bevington and Robinson 1969):

$$\mu = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3}{w_1 + w_2 + w_3} \qquad \text{with } w_i = \frac{1}{{\sigma_i}^2} \qquad \text{for } i = 1,\, 2,\, 3 \tag{5.1}$$

$$\mu = \frac{(x_1/{\sigma_1}^2) + (x_2/{\sigma_2}^2) + (x_3/{\sigma_3}^2)}{(1/{\sigma_1}^2) + (1/{\sigma_2}^2) + (1/{\sigma_3}^2)} \tag{5.2}$$

Assuming that the fluctuations in the x_i's are uncorrelated, the variance of the weighted average is the inverse of the sum of the weights:

$${\sigma_\mu}^2 \simeq {\sigma_1}^2 \left(\frac{\partial \mu}{\partial x_1}\right)^2 + {\sigma_2}^2 \left(\frac{\partial \mu}{\partial x_2}\right)^2 + {\sigma_3}^2 \left(\frac{\partial \mu}{\partial x_3}\right)^2 \tag{5.3}$$

$${\sigma_\mu}^2 \simeq \frac{1}{(1/{\sigma_1}^2) + (1/{\sigma_2}^2) + (1/{\sigma_3}^2)} \tag{5.4}$$

These formulas could not be applied directly to the data, however, because the flux levels were not the same in all three spectra. For most orders, the fluxes in casp0031.bdf were higher than in the other two spectra. In an attempt to correct for this effect, the three median values were used to scale the casp0023 and casp0089 spectra to the level of the casp0031 spectrum before all three were averaged.

So the median value m_i of each order spectrum was used to scale that spectrum by a factor of $c_i = m_1/m_i$. The weighted average of the *scaled* spectra thus became

$$\mu = \frac{w_1(c_1x_1) + w_2(c_2x_2) + w_3(c_3x_3)}{w_1 + w_2 + w_3} \quad \text{with} \quad w_i = \frac{1}{{c_i}^2{\sigma_i}^2} \tag{5.5}$$

$$\mu = \frac{x_1/(c_1{\sigma_1}^2) + x_2/(c_2{\sigma_2}^2) + x_3/(c_3{\sigma_3}^2)}{1/({c_1}^2{\sigma_1}^2) + 1/({c_2}^2{\sigma_2}^2) + 1/({c_3}^2{\sigma_3}^2)} \tag{5.6}$$

with a variance of

$${\sigma_\mu}^2 \simeq \frac{1}{1/(c_1\sigma_1)^2 + 1/(c_2\sigma_2)^2 + 1/(c_3\sigma_3)^2} \tag{5.7}$$

After each of the orders were combined according to equation (5.6), they were kept separate at this point in case order-by-order corrections might later become necessary.

Of course it would have been possible to merge the orders from each QSO exposure into a full spectrum first, and then combine the full spectra from the three exposures. But it would have been much more difficult to scale the full spectra appropriately. This was because the flux levels differed not only from one exposure to another, as seen above, but also within the different wavelength regions of each exposure. For example, the second QSO exposure had slightly more flux at lower wavelengths, but clearly less flux at higher wavelengths, than the third QSO exposure.

5.2 Order Merging (II)

Normally, the next step would have been to estimate the continuum in each of the orders, then to normalize the orders and merge them to produce the complete spectrum. However, when it was noticed that the measured flux did not drop to zero in the middle of obviously saturated lines, it was clear that not all of the sky background had been effectively subtracted and that a sky-subtraction step had to be added.

In order to enable a global estimate of the sky background, the 43 combined orders had to be merged into a complete spectrum at this point. This was not such a simple matter, however. Since `MERGE/ECHELLE` requires a rebinned frame as input, the procedure necmerge.prg and the source-code file necmerge.for (MIDAS release 95NOV) were modified so that many separate order frames would be accepted instead. The second, more important difficulty was that the `MERGE/ECHELLE AVERAGE` algorithm had to be adapted to enable it to merge the error frames as well as the data frames. The fluxes could be averaged as described in section 4.7:

$$\mu = w_1 x_1 + w_2 x_2, \qquad \text{with } w_1 + w_2 = 1 \tag{5.8}$$

The corresponding errors, on the other hand, could not be averaged in the same way as the fluxes. According to the rules of error propagation (cf. equation 5.3), the error σ_μ for the weighted average is given by

$${\sigma_\mu}^2 = {w_1}^2 {x_1}^2 + {w_2}^2 {x_2}^2 \tag{5.9}$$

That is, the weights used for the variances must be the *squares* of the weights used for the data.

The modified versions of necmerge.prg and necmerge.for were therefore given an additional input parameter to specify whether the pixel values being averaged were fluxes or variances; the modified algorithm assigned the normalized weights from 0 to 1 accordingly. Thus the full spectrum (Fig. 5.1) resulted from the equivalent of `MERGE/ECHELLE AVERAGE`, while the corresponding error frame was obtained by squaring all of the individual order frames, merging them using the variance option of the modified procedure, and then taking the square root of the resulting output frame.

5.3 Heliocentric and Vacuum Corrections

It is standard practice to correct the wavelengths in a spectrum for the Earth's orbital motion along the line of sight, as well as to convert the measured wavelengths to their vacuum values. These corrections were made for the 43 individual order frames as well as for the merged spectrum described in section 5.2.

An object's heliocentric velocity v_H (i.e., with respect to the Sun) is determined by removing the velocity component resulting from the Earth's orbital motion. For an object with right ascension α_o and declination δ_o, the heliocentric radial velocity is related to its geocentric velocity v_G (i.e., measured from the center of the Earth) by

$$v_H = v_G + v_E[\sin\delta_A \sin\delta_o + \cos\delta_A \cos\delta_o(\alpha_o - \alpha_A)] \tag{5.10}$$

where v_E is the Earth's orbital velocity (29.78 km s^{-1} on average),

$$\alpha_A = \arcsin(\cot\delta_s \sin\alpha_s \sin\varepsilon) \tag{5.11}$$

$$\delta_A = \alpha_s - \arccos(-\tan\delta_A \tan\delta_s) \tag{5.12}$$

are the right ascension and declination of the apex of the Earth's orbital motion at the time of the observation, ε is the obliquity of the ecliptic ($23°27'$), and α_s and δ_s are Sun's right ascension and declination at the time of the observation (Kitchin 1995). Clearly, the difference between the geocentric and heliocentric velocities—which is the

Earth's heliocentric radial velocity—causes a shift $\Delta\lambda$ in the wavelength according to the Doppler effect,

$$\frac{\Delta\lambda}{\lambda} = -\frac{\Delta\nu}{\nu} = \frac{v}{c} \tag{5.13}$$

where c is the speed of light.

The calculations were done automatically with the MIDAS commands `COMPUTE/PRECESSION`, which precessed the celestial coordinates of HE 0940 (listed in Table 2.1) to the day of observation, followed by the command `COMPUTE/BARYCORR`, which computed the Earth's heliocentric radial velocity at the desired time. The total heliocentric correction to the radial velocity ranged between 7.6 km s^{-1} and 8.7 km s^{-1}, which led to a wavelength correction of approximately 0.1 angstroms.

Converting the wavelengths to their vacuum values is a larger correction that requires an expression for the index of refraction n of the Earth's atmosphere. For convenience, an estimate was used that applies to optical wavelengths (2000–7000Å) and is independent of the time and location of the exposure:

$$(n-1) \times 10^7 = 643.28 + \frac{294981}{146-\sigma^2} + \frac{2554.0}{41-\sigma^2} \tag{5.14}$$

where σ is the wave number in vacuum ($1/\lambda_{\rm vac}$), in units of reciprocal microns (Lang 1980). The vacuum wavelength is merely

$$\lambda_{\rm vac} = n\ \lambda_{\rm air} \tag{5.15}$$

In the case of the HE 0940 spectra, the correction $\Delta\lambda = \lambda_{\rm vac} - \lambda_{\rm air}$ resulted in shifts between 1.07 and 1.47 Å, also comparable to the approximate value $n \approx 1.000297$ that is often used (Lang 1980).

On a technical note, the preceding corrections represented a change in the wavelength axis of the spectra and resulted in a non-uniform wavelength step from one pixel to the next. But MIDAS frames currently allow only a single step size. So during these corrections, a table was created to hold the wavelengths, fluxes and errors from each input frame and its corresponding error frame. The data were subsequently kept in that form, and all subsequent steps were performed on these tables.

5.4 Sky Subtraction

Because of the small slit size used for the object exposures, the quasar most probably filled the entire slit. In any case, the spatial profile along the slit was examined before and during the spectrum extraction and no sky background had been recognizable; so the sky background could not be extracted separately. There was no other choice but to estimate the level of the sky background at this point by examining the line centers of saturated lines.

It was assumed that the sky background would be a slowly varying function of the wavelength. A global fit was therefore thought to be preferable to order-to-order fits, which would not only necessitate smoothing of discontinuous fits in the overlaps between two orders but probably lack sufficient data points in some of the orders anyway. That is why the combined orders were merged as described in section 5.2, even though the continuum definition was later to be done on each of the orders individually.

Points at the center of lines judged to be obviously saturated were selected manually. The sky background level was estimated by fitting a fourth-degree polynomial through these 492 out of a total of 11,710 data points. The standard deviation of the fit was 0.45, and the resulting sky level ranged between 0.67 and 2.54, in the arbitrary flux units of the spectrum shown in Figure 5.1.

Since there were no saturated lines redward of the line at 4910 Å, there was no way to know how the sky background should be extrapolated to higher wavelengths. The fourth-degree polynomial fit was used up to 4910.3 Å, and the fit value there (1.9557) was used for all wavelengths beyond that point. This value is doubtless a bit high in the area of the C IV λ1548 lines around 5155 Å; six pixel values fall below the estimated sky there. But in the absence of any information about the sky level, it was thought best to use the constant value rather than making any further assumptions about the shape of the sky spectrum or allowing a discontinuity by assuming no sky background above 4910.3 Å. In the end, this was also the estimate that was subtracted from all of the flux values in the individual orders, just before a fit of the continuum was attempted.

5.5 Spectrum Normalization

The next step consisted of normalizing the spectrum, i.e., scaling the flux so that the pixels in which there was no absorption would have pixel values that fluctuate (at levels consistent with the noise there) about a constant level of 1.

In addition to the reasons mentioned in section 5.1, the orders were first combined (rather than merged) because it proved much easier to estimate the continuum for each of the 43 echelle orders individually. This also allowed more flexibility in fitting the edges of the orders, where the lower fluxes meant that the background subtraction and response correction had more strongly affected the determination of the flux than in the center of the order.

The most common method of estimating the continuum in echelle spectra is to begin by fitting polynomials to the spectrum and discarding the pixel values that differ more than a certain number of standard deviations from the fit. The process is then repeated iteratively, usually until the resulting residuals correspond to levels consistent with the expected noise. The aim is to remove from consideration any pixels

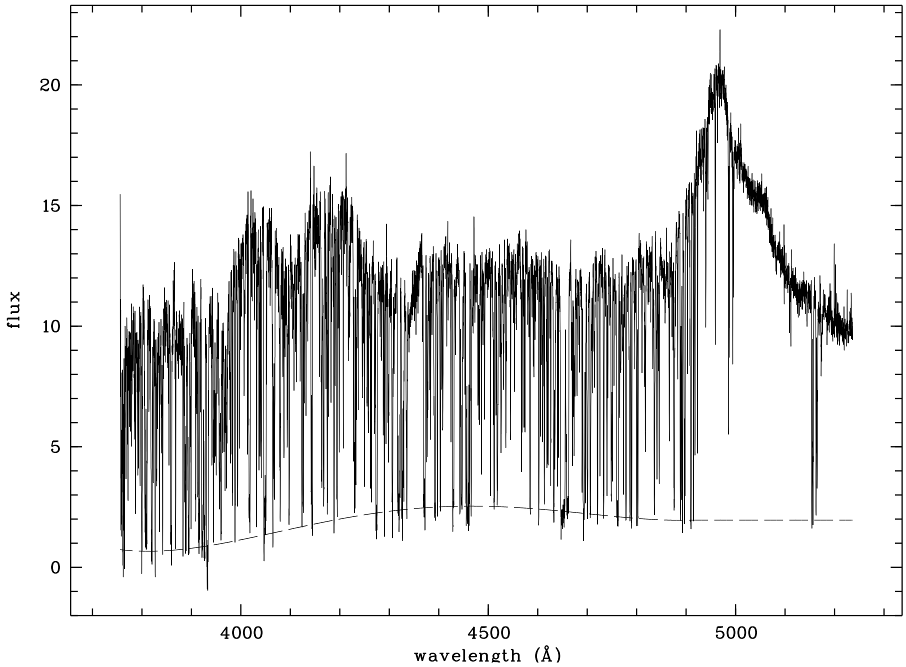

Figure 5.1: *Preliminary merged spectrum with estimated sky background.*

belonging to absorption or emission lines, noise spikes, or cosmics before performing the final fit that represents the estimated continuum.

Strictly speaking, normalizing the flux values in the spectrum is not quite the same as estimating the continuum, because a continuum fit in the above sense would aim to ignore emission lines as well as absorption lines. When a spectrum is normalized, however, any emission lines that might have been present are fit along with the continuum to produce the absorption spectrum, so only pixel values below a certain cut-off would be excluded from the fit. Nevertheless, such a fit is also referred to as the estimated continuum, and the normalized spectrum is created by dividing the flux by this continuum.

In this particular case, no automated fitting procedure was used. Instead, points in each echelle order were first selected as belonging to the continuum and then used to determine the fit. The principal difficulty lay of course in making the subjective judgment as to which pixels were unaffected by absorption. Under the necessary assumption that at least some points in each order did belong the continuum, an attempt was made to select all of the pixels that did not appear to be part of absorption lines. Groups of adjacent pixels were selected rather than only the very highest values, to account for the fluctuations due to noise. The merged spectrum obtained in section 5.2 was also consulted during the selection process since it provided some visual guidance as to the global continuum level.

The selection and fitting were made more difficult by the fact that the large number of absorption lines in most of the orders left very little continuum available, so that usually only very small groups of points could be chosen. The number of values used in each fit reached a maximum of 514 in order 39, and included well over 60% of the pixels in all of the other echelle orders redward of Lyman α emission. The number of fit points dropped to a minimum of only 24 pixels in order 7 (see Table 5.1); but such low values are no surprise in orders 1 to 15, which lie blueward of Lyman β emission and thus contain additional lines from Ly β and even Ly γ absorption. In most cases within the primary region of the Ly α forest, however, between 80 and 100 values were used in each fit. The continuum is estimated to cover between 15% to 20% of the spectrum in most orders. The markedly higher percentages in orders 20 and 21 are due to the presence of a possible void, which will be investigated in chapter 9.

Once the input values had been selected, second- or third-degree polynomial fits were then performed on the chosen pixels. Low-degree fits were purposely used because some orders lacked fit points on one or both edges—and in such cases higher-degree fits tend to fluctuate widely, curving sharply upward or downward depending on the degree. Only in three echelle orders did polynomials of degree 4 or 5 seem necessary.

It should be noted in passing that since this type of method assumes that the highest pixel values in each order do indeed represent continuum levels, it will tend to consistently underestimate the continuum level by ignoring smoothly fluctuating absorption such as that from weak, very broad lines or, on a cosmological scale, from diffuse intergalactic hydrogen (the Gunn-Peterson effect). Of course it would be difficult to argue that the interactive fitting was able to correct more for this bias than automated fitting. The latter would probably have yielded more consistent, or at least more quantifiable, results from one order to the next. Admittedly, however, the continuum fitting was done by hand less for consistency than for educational purposes—to gain first-hand experience with the problems inherent in continuum estimates for spectra so densely populated with absorption lines.

5.6 Order Merging (III)

The normalized orders and their correspoding errors were now merged using the same algorithm as described in section 5.2, although some changes were necessary now that the input was present in the form of tables rather than one-dimensional images. The algorithm was also modified further in order to obtain the greatest possible wavelength coverage in the merged spectrum.

When the MIDAS command `MERGE/ECHELLE AVERAGE` merges two echelle orders, it assumes that the overlap between them is bounded by the minimum of the redder order on the low-wavelength end and by the maximum of the bluer order on the high-wavelength end. This is usually the case, so usually the interval whose length corresponds to the echelle keyword `DELTA` is excluded from the appropriate end of each order. This exclusion feature was retained in the modified procedures because despite normalization, the flux values on the ends of each order showed occasional spikes and were generally not very reliable. An interval of 5 Å was chosen to avoid these regions without sacrificing the good overlap already present.

Although the Hamburg procedures had allowed the optimal extraction of a greater spectral region than would have been possible otherwise, only a small part of order 43 was present and its maximum wavelength was less than the minimum wavelength in order 42. In this case, the normal MIDAS algorithm would have ignored the highest wavelengths in order 42 and kept the much less reliable values at the end of order 43. The modified algorithm, on the other hand, first checked the positions of the order boundaries when deciding at which ends to exclude values before averaging the overlaps. The final normalized and merged spectrum of HE 0940 was obtained via this modified algorithm.

Relat. order	Echelle order	Minimum wavelength	Maximum wavelength	Data values	Contin. values	Contin. coverage	poly. degree
1	150	3755.98	3808.70	417	71	17 %	3
2	149	3781.20	3834.17	419	94	22 %	2
3	148	3806.80	3860.03	421	90	21 %	3
4	147	3832.53	3886.26	425	61	14 %	3
5	146	3858.76	3913.13	430	52	12 %	2
6	145	3885.37	3939.87	431	40	9 %	3
7	144	3912.37	3967.37	435	24	6 %	2
8	143	3939.74	3994.00	437	54	12 %	2
9	142	3967.37	4023.13	441	123	28 %	2
10	141	3995.50	4051.65	444	154	35 %	5
11	140	4024.02	4080.54	447	71	16 %	2
12	139	4052.92	4109.82	450	57	13 %	2
13	138	4082.32	4139.73	454	47	10 %	2
14	137	4112.10	4169.89	457	78	17 %	2
15	136	4142.14	4200.56	462	101	22 %	2
16	135	4172.93	4231.61	464	133	29 %	2
17	134	4203.98	4263.04	467	93	20 %	3
18	133	4235.54	4295.10	471	93	20 %	2
19	132	4267.60	4327.67	475	67	14 %	4
20	131	4300.17	4360.62	478	229	48 %	3
21	130	4333.25	4394.21	482	217	45 %	4
22	129	4366.83	4428.30	486	83	17 %	2
23	128	4400.80	4462.77	490	90	18 %	2
24	127	4435.40	4497.88	494	98	20 %	2
25	126	4470.63	4533.49	497	148	30 %	2
26	125	4506.37	4569.73	501	81	16 %	2
27	124	4542.74	4606.61	505	92	18 %	2
28	123	4579.62	4644.13	510	90	18 %	2
29	122	4617.00	4682.02	514	77	15 %	2
30	121	4655.28	4720.67	517	74	14 %	2
31	120	4693.93	4760.09	523	55	11 %	2
32	119	4733.35	4800.01	527	52	10 %	2
33	118	4773.39	4840.69	532	108	20 %	2
34	117	4814.08	4881.88	536	85	16 %	3
35	116	4855.64	4923.95	540	103	19 %	2
36	115	4897.85	4966.79	545	284	52 %	3
37	114	4940.68	5010.39	551	384	70 %	3
38	113	4984.41	5054.62	555	393	71 %	2
39	112	5028.76	5099.61	560	514	92 %	2
40	111	5074.01	5145.61	566	475	84 %	2
41	110	5120.14	5192.38	571	362	63 %	2
42	109	5167.03	5239.90	576	374	65 %	3
43	108	5214.81	5236.35	171	171	100 %	2

Table 5.1: *Echelle orders and continuum fits*