

Chapter 8

Data reduction

Figure 8.1 shows a typical spectrum as it was obtained during the observation campaign. On the left side is the raw spectrum before data reduction, the right side shows the fully reduced absolute flux calibrated spectrum. The contribution from reflected sun light on the cometary dust particles was not removed. The labeling of the axes shows the nomenclature used for the following sections. For not reduced or only partly reduced spectra the axes are simply rows and columns on the CCD. Only after full calibration these values can be linked to the physical properties: projected nucleocentric distance ρ and wavelength λ .

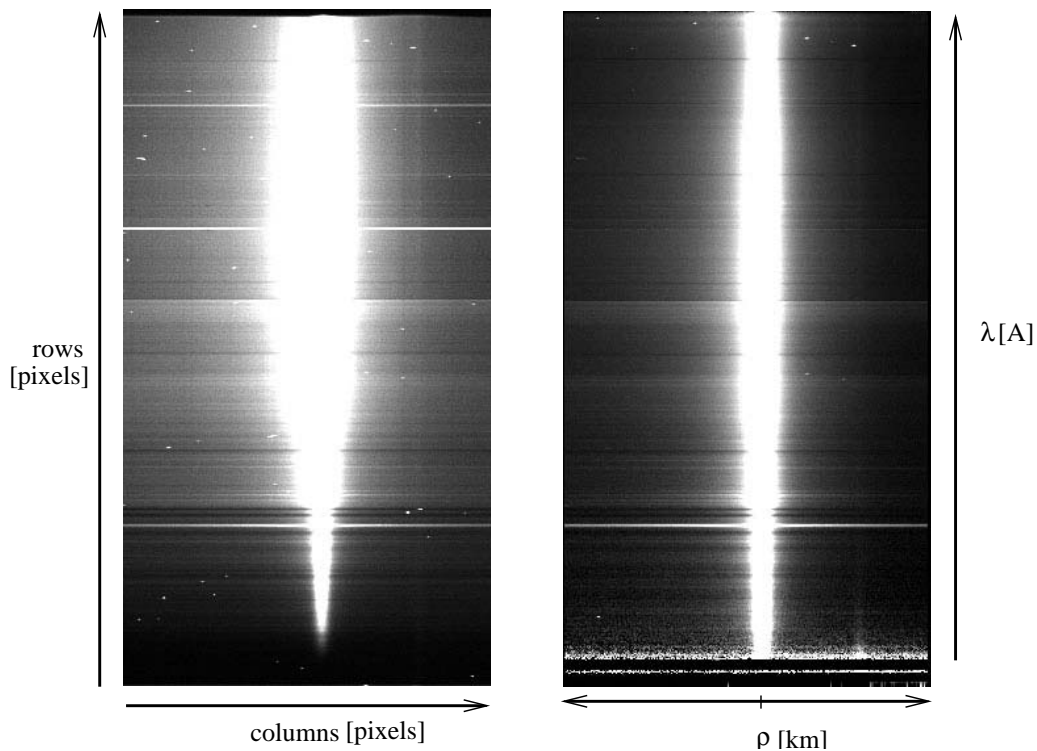


Figure 8.1: The left panel shows a longslit spectrum before data reduction, the right panel shows the same spectrum fully reduced.

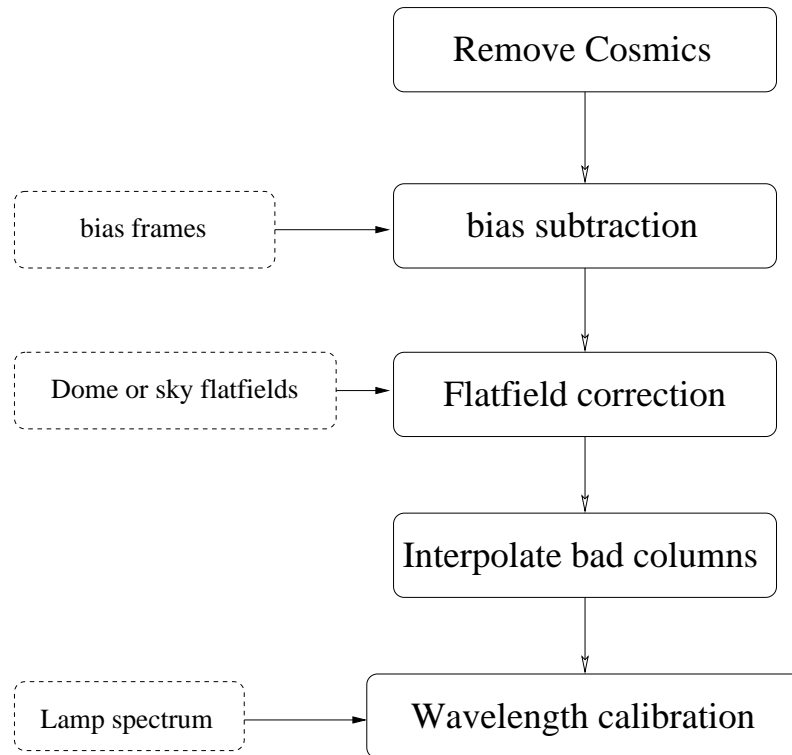


Figure 8.2: Flowchart for first part of the data reduction

The following sections will give an overview of the data reduction process. The software used for the reduction follows the standard CCD reduction scheme and has been updated for this work based on existing routines. Generally the data reduction contains two parts, the first part as shown in Figure 8.2 deals mainly with instrumental effects, while the second part shown in Figure 8.3 deals with environmental effects.

Remove cosmics

CCDs are not only sensitive to visible light. Radiation from cosmic rays or natural background radioactivity can also create a signal on the CCD. These signals, usually very bright spots or streaks, depending on the incident angle of the radiation, are often called 'cosmics'. On a spectrum cosmics have to be identified and removed. The identification of a cosmic is a tedious task. The best way to do this is manually. The human brain is a very efficient filter to distinguish between 'real' emission features and superimposed signals. However, the sheer amount of data makes a semi-automatic approach necessary. Each spectrum is first scanned by an automatic routine. This routine marks all signals which are above a certain intensity compared to its surrounding as a possible cosmic. The user can change the threshold as well as the radius for the surrounding. All identified cosmics are removed from the spectrum by replacing the affect area with an interpolation in spatial direction. Depending on the quality of the spectrum this first step removes about 70-80% of all cosmics. The remaining cosmics are marked and removed manually in a second step. Removing the cosmics is a difficult

task, because it has to be done without affecting any (or as little as possible) of the very faint emission features of the comet. Once the cosmic is identified by eye the affected area is again replaced by an interpolation in spatial direction. Using the spatial direction for the interpolation is important, because in an emission feature this direction shows a steady gradient with a generally shallow slope. In wavelength direction the gradient can be very steep especially for narrow emission bands, e.g. the CN (0-0) emission.

Bias

The digital to analog (D/A) converter of the CCD, introduces an offset to avoid negative digital numbers. In order to quantify this offset a number of bias frames are obtained during the night. The bias level depends on a number of parameters, including the temperature the electronics and power level changes and can vary over the night. These variations can be detected by plotting the average bias level for each bias frame versus time. If these bias level shows little variation (smaller than 5%) a median bias frame is created from all bias frames of the night. In the case of variations during the night, stable periods are identified and for each period a separate mean bias frame is generated. If the bias level is constant or nearly constant the mean bias frame is replaced by the mean bias level (a single number instead of an always noisy frame). This is subtracted from each pixel in all other frames. Should the bias level show variations over the CCD, the mean bias frame is subtracted from each other frame. On some CCDs there is an area on the edges which is covered and does therefore receive no light. This area is called the oversampled margin. Since this margin receives no light, the signal is solely due to the bias level. With this information a bias level can be obtained from each frame. The bias level from each frame is used to scale a normalized mean bias, which is then subtracted from the frame. This ensures that the spatial variation over the CCD and the (small) temporal variations over the night are taken into account in an optimal way.

For the VLT/UT1 observations so called 'Master bias' frames are provided by ESO as part of the standard calibration process. These 'Master bias' are already processed. After carefully checking these images, the best matching 'Master bias' frame has been used for the calibration. The match was mainly determined by the residuals in the oversampled margin of the data frames.

Flatfields

A CCD shows a pixel to pixel variation of the sensitivity. Furthermore the slit and the optical systems are not perfect, leading to spatial differences in the transmission properties. To quantify these flatfield images are obtained for each set up used during the observation. A flatfield image is obtained by pointing the telescope at a homogeneously illuminated surface, usually a screen in the dome. With each setup a series of flatfield images is made. From these images a median flatfield image is obtained for each setup. The median flatfield is normalized and each frame obtained with this setup is divided by the normalized median flatfield frame.

Again for the VLT/UT1 observations so called 'Master flatfields' are provided by ESO as part of the standard calibration process. These 'Master flatfield' are already processed and normalized. After carefully checking these images, they have been used for the calibration. In the case of observations obtained at the Danish 1.5m telescope, the straylight reflex observable in the blue part of the spectrum could not be fully removed by the flatfield procedure, since the reflex was not in a fixed location on the CCD frame. Furthermore it varies in intensity, both absolute and relative.

Interpolate bad pixel/columns

Some CCD's show defects. The most common defect are columns of pixels, which are either always saturated or which are always on zero. In order to eliminate these columns a procedure is used which replaces the affected pixel in each column with a value determined from a spline-fit to the surrounding pixels in every row.

Wavelengths calibration

The dispersion function of a grism is measured during the night by observing calibration lamps. Standard calibration lamps are Helium, Argon or Neon lamps. The dispersion function is determined by a polynomial fit, giving the relation between the pixel position on the CCD and the wavelengths. Using this dispersion curve all spectra are calibrated. The rows of each spectrum now correspond to a defined wavelength. Whenever possible, additional lamp spectra of the same night have been used to check the wavelength calibration.

Extinction

Light from the comet or a star has to pass through Earth's atmosphere. Atmospheric extinction leads to a wavelength dependent attenuation of the light. The difference between the observed object brightness and the true brightness, expressed in magnitudes, can be defined as a factor

$$f_{\text{ext}}(\lambda) = A \cdot \gamma(\lambda) \quad (8.1)$$

$\gamma(\lambda)$ is the extinction function. The airmass A is a measure of the height of atmosphere through which an observation is being made.

For each telescope location a standard extinction function $\gamma_{\text{standard}}(\lambda)$ has been determined. This curve is given in a tabulated form in the documentation for the telescope location (for La Silla for example in ESO User Manual 1993 [ESO, 1993] and Burki *et al.* [1995]). Converting equation (8.1) from magnitudes to fluxes a correction factor $f_{\text{ext}}(\lambda)$ can be derived

$$f_{\text{ext}}(\lambda) = 10^{0.4 \cdot \gamma(\lambda) \cdot A} \quad (8.2)$$

The correction factor $f_{\text{ext}}(\lambda)$ is applied to each spectrum $F(\lambda)$

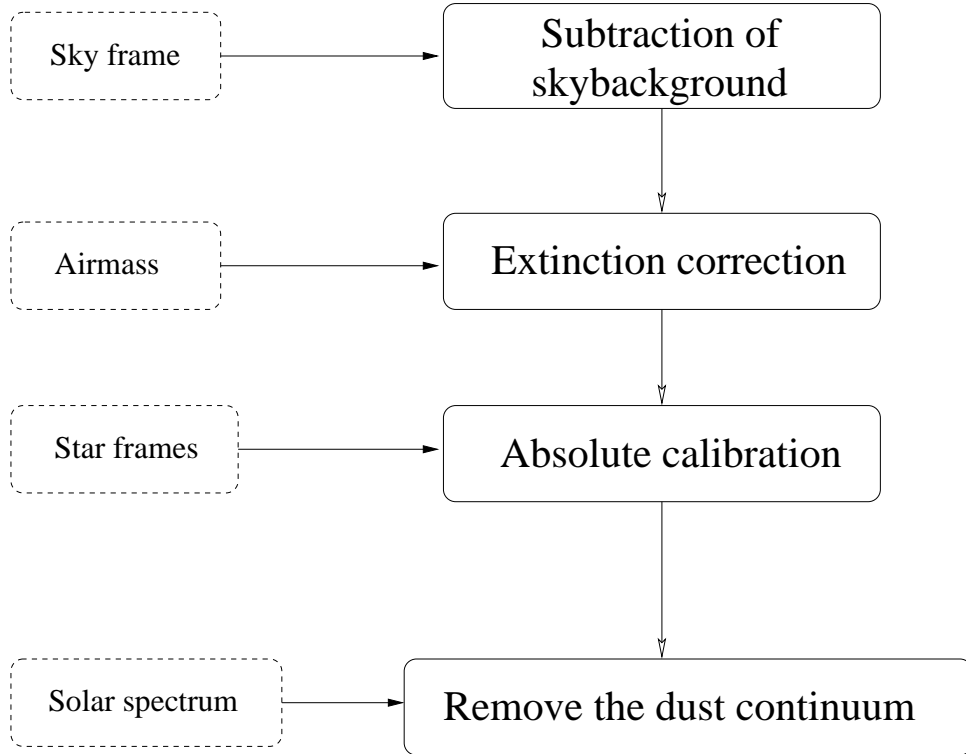


Figure 8.3: Flowchart for the second part of the data reduction

$$F_e(\lambda) = F(\lambda) \cdot f_{\text{ext}}(\lambda) \quad (8.3)$$

For observations at the VLT/UT1 and the ESO 3.6m telescope the airmass A is given in the header of each spectrum. The values are calculated by ESO at the time of observation using `skycalc` by Thorstensen [1993].

For the observations at other telescopes the values for A have been calculated based on the position of the object on the sky using the approximation for a plane parallel atmosphere

$$A = [\sin \phi \sin \delta + \cos \phi \cos \delta \cos h]^{-1} \quad (8.4)$$

- ϕ latitude of the observatory
- δ declination of the observed object
- h hour angle

Absolute calibration

The sensitivity of the telescope system is wavelengths dependent. CCD's are usually more sensitive in the red wavelength range of the optical spectrum than in the blue. The transmission and reflection properties of the optical elements in the light path of the telescope

are wavelengths dependent as well. These effects combined can be described by the response function $k(\lambda)$ with

$$F_{\text{measured}}(\lambda) = k(\lambda) \cdot F(\lambda) \quad (8.5)$$

Here $F_{\text{measured}}(\lambda)$ is the flux as measured by the CCD and $F(\lambda)$ is the flux emitted by the object.

Up to this reduction step the flux data are in 'data units' (DU). These units have to be converted to physical units for further analysis.

The response function has to be obtained for each night and for each optical setup. For this purpose a spectrophotometric standard star is observed. A list of all standard stars used for the calibration is given in Table C.1 in Appendix C. For these stars the spectra F_{star} in physical units are given in the literature (for example Oke [1990], Turnshek *et al.* [1990] and Bohlin *et al.* [1990] for Hubble Space Telescope calibration stars, Taylor [1984] and Hamuy *et al.* [1994], Hamuy *et al.* [1992]).

While the spectrophotometric standard star spectrum F_{star} is given in units of [$\text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1}$], the observed spectrophotometric standard star spectrum F_{star}^* is measured in units of [DU pix^{-1}].

In a first step both spectra are rebinned to the same wavelength resolution, a 50 \AA wavelength bin.

$$\begin{aligned} F_{\text{Star}}^{\text{bin}} &= \frac{\sum_i F_i}{\text{bin}} \\ F_{\text{Star}}^{*\text{bin}} &= \frac{\sum_i F_i^*}{\text{bin}} \end{aligned} \quad (8.6)$$

where $\text{bin}=50 \text{\AA}$ and F_i and F_i^* are the fluxes in the original bins. $F_{\text{Star}}^{\text{bin}}$ is still a flux per \AA , but now in the middle of a 50 \AA bin and the same is true for $F_{\text{Star}}^{*\text{bin}}$.

The response function k of the instrument system is yielded by the ratio of spectrum as given in the literature to the observed spectrum

$$k^{\text{bin}} = \frac{F_{\text{Star}}^{*\text{bin}}}{F_{\text{Star}}^{\text{bin}}} \quad (8.7)$$

k^{bin} is in units of [DU pix^{-1}]/[$\text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1}$]. k still has a 50 \AA wavelength resolution. k^{bin} is then interpolated to the wavelength resolution $(n \text{\AA})^{-1}$ of the cometary spectra by a linear interpolation yielding the response curve k . The absolute calibration of the cometary spectrum F_{Comet} is then performed by

$$F_{\text{Comet}}^* = \frac{F_{\text{Comet}}}{k} \left[\frac{\text{DU} \text{\AA}^{-1}}{\text{DU} \text{\AA}^{-1}} \text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1} \right] = \left[\text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1} \right] \quad (8.8)$$

There is a caveat here in the fact, that the flux in the cometary spectrum is in units of \AA^{-1} but the wavelength resolution is in general not 1\AA . This means that the flux is given for the middle of a bin of width $n\text{\AA}$, which is the actual wavelength resolution. In other words the spectrum is sampled in $n \text{\AA}$ steps.

Check of error in calibration

As can be seen in Table C.1 in Appendix C for some nights more than one spectrophotometric standard star has been observed. All additional standard star spectra were calibrated using the same response function k . The flux of the observed calibrated spectrum F_{calib} can then be compared to the flux given in the standard star catalogs F_{cat} . The mean ratio

$$R = \frac{\overline{F_{\text{calib}}}}{F_{\text{cat}}} \quad (8.9)$$

is a measure for the quality of the calibration and for the stability of the observing conditions. The mean is calculated over the wavelength range from $4500\text{--}5500 \text{\AA}$. This sub range was chosen, because all observations on the different telescopes and with the different setups cover this range and it is not contaminated by the reflex in the blue wavelengths range at the 1.5m Danish telescope.

$$Q = |1 - R| \cdot 100. \quad (8.10)$$

yields the relative error in percent of the calibration within a night.

The calibration over the whole observing period of nearly 5 years has been checked. Unfortunately a large number of different spectrophotometric standard stars have been used during this observing period. This makes it nearly impossible to compare the absolute calibration directly by a ratio as defined in equation (8.9). Lets assume in the first night the spectrophotometric standard star α was observed yielding the absolute calibrated spectrum $F_{\text{Obs}\alpha}$. In the second night the standard star β was observed resulting in the absolute calibrated spectrum $F_{\text{Obs}\beta}$. With the spectra $F_{\text{Cat}\alpha}$ and $F_{\text{Cat}\beta}$ given in the literature for the stars α and β the ratios

$$R_1 = \frac{F_{\text{Obs}\alpha}}{F_{\text{Cat}\alpha}} \quad (8.11)$$

$$R_2 = \frac{F_{\text{Obs}\beta}}{F_{\text{Cat}\beta}} \quad (8.12)$$

$$(8.13)$$

for the first and the second night can be determined. These two ratios yield the relative calibration of the nights by

$$R_{\text{inter}} = \frac{R_1}{R_2} \quad (8.14)$$

The reasoning for equation (8.14) is obvious, if one assumes that the standard stars α and β are identical. The relative error Q_{inter} as a percentage is given by

$$Q_{\text{inter}} = |1 - R_{\text{inter}}| \cdot 100. \quad (8.15)$$

The spectrophotometric standard stars used for the absolute calibration and the quality of the calibration within a night Q as well as the quality of the calibration over the whole range of observed nights, the inter-calibration quality Q_{inter} , are given in table C.1 in appendix C. As a reference the night of the 21/22.1.1998 was used. This night was a night with good observing conditions and the quality of the calibration within the night Q (as defined in equation (8.10)) was the best of all nights. Since the value Q_{inter} gives only a relative error, the values can easily be recalculated using a different night as reference point.

Solar continuum

The observed spectrum of a comet is a combination of the emission spectrum of the gas coma and the reflection spectrum of the dust coma.

$$F_{\text{Comet}} = F_{\text{Gas}} + F_{\text{Dust}} \quad (8.16)$$

The dust spectrum F_{Dust} can be written as

$$F_{\text{Dust}} = C(\lambda) \cdot F_{\odot}$$

F_{\odot} is the solar spectrum and $C(\lambda)$ is a function describing the reflectivity and scattering of the dust.

In order to extract the gas emission spectrum F_{Gas} from the observed cometary spectrum F_{Comet} the function $C(\lambda)$ has to be determined.

There are two ways to obtain the solar spectrum F_{\odot} . The best way would be to observe the solar spectrum directly with the same instrumental setup as the comet. Of course this is very difficult due to the brightness of the Sun. It is possible to observe the solar light reflected from the Moon. Still this target might be too bright. A practical solution is to observe a solar analog star. These stars are of the same spectral class as our Sun and have therefore a nearly identical spectrum.

If the solar spectrum has not been observed in the same night, the solar spectral atlas by Kurucz *et al.* [1984] can be used. The solar spectrum given here has a much higher wavelength resolution ($<800\text{nm}$ in 0.0005 nm bins and for $\geq 800\text{ nm}$ in 0.001 nm bins) than the observed spectra. For this reason it is convolved with a Gaussian profile to degrade the resolution.

Solar analog stars have only been observed in some observing runs at the Danish 1.52m telescope. For this study for all nights the solar spectrum from Kurucz *et al.* [1984] has been used.

The convolved solar spectrum is compared with the cometary spectrum. Solar absorption bands have to be identical in width and depth.

The approach chosen in this work to obtain $C(\lambda)$ is to approximate the function by a polynomial $C(\lambda)$ of fifth degree.

The coefficients $a_0 \dots a_5$ are determined by fitting the solar spectrum F_{\odot} to the cometary spectrum. For this step the spatial resolution of the spectra is reduced by averaging over 10 to 20 pixels depending on the quality of the spectrum to increase the signal to noise ratio. In each of the resulting columns six wavelengths regions are defined which appear to be free of gas emissions. The polynomial $C(\lambda)$ is determined by a polynomial fit to these regions. The spectrum $C(\lambda) \cdot F_{\odot}$ is then subtracted from the cometary spectrum. The procedure is repeated for each column, resulting in the emission spectrum F_{Gas} of the gas. As a byproduct $C(\lambda) \cdot F_{\odot}$ from each column yields the calibrated dust spectrum F_{Dust} .

For very noisy spectra or for the contaminated spectra obtained at the 1.52m Danish telescope, the spectrum is divided in subspectra in the wavelength direction. Each of these subspectra is then fitted separately.

For the observations at the VLT/UT1 at heliocentric distances greater than 9.8 AU only the spectral range from 3700 to 4300 Å was treated separately. This allowed to subtract the solar continuum on this wavelengths interval very accurately.

Integration of emission lines and bands

Two approaches have been used to obtain the integrated flux within an emission band. For the emission bands of C₂ and C₃ the area of the emission band is integrated by

$$F_{\text{int}} = \sum_i F_i \cdot d\lambda \quad (8.17)$$

where i covers all bins within the emission band and $d\lambda$ is the wavelengths bin.

For the CN and the NH₂ emissions a different approach is used. CN has a very narrow band which appears in low and medium resolution spectra as a single emission peak structure. The NH₂ emissions show two narrow bands which appear as a double peak structure. The internal structure of these emission bands are not resolved by the medium resolution spectra used in these observations. Each peak is fitted with a Gaussian profile plus a constant background. The Gaussian profile is a first order approximation of the band structure. The area under the profile is integrated numerically using Simpsons rule. For NH₂ the results for the two peaks are added. Approximating the real band structure by a Gaussian profile can introduce a small additional error. However visual inspection of the profiles fitted to the band show very little differences.

There are two further sources of errors which have been taken into account. These are the so-called 'guiding' error and the so-called 'continuum' error.

The 'guiding' error is due to changes in the positioning of the slit relative to the comet position. The guiding at a telescope is never perfect. During a series of exposures it can trail off slightly, thus the name 'guiding' error. The off center guiding results in a variation of the illumination of the slit. To quantify this effect, an error can be estimated from the continuum spectra. A part of the continuum spectrum is identified which is free of residuals of emission lines of the comet or solar absorption lines. The selected wavelengths range in

each spectrum is averaged in wavelength direction to obtain a spatial distribution of the brightness. For a perfectly guiding telescope and without any changes in the atmospheric seeing, the difference between the spatial profiles obtained from separate spectra should be almost zero. Thus any difference is a direct measure for the 'guiding' error.

$$e_{\text{guiding}}(\rho) = \frac{s(\rho)}{m(\rho)} \quad (8.18)$$

where $s(\rho)$ is the standard deviation of the profiles and $m(\rho)$ is the mean of the profiles over nucleocentric distance.

The 'continuum' error occurs only for the integration method used in 8.17 in case of an over- or under-subtraction of the solar continuum. To determine this error quantitatively, two subspectra F^{left} and F^{right} in the cometary spectra are selected, bracketing the emission band. These subspectra must be free of emission features and they should ideally have the same width in wavelength, to guarantee an equal weighting of both sides. Both subspectra are then averaged in wavelength direction. The resulting spatial profiles should be zero, if the subtraction of the solar continuum was done perfectly. Any difference is a direct measure of the error. Multiplication with the width of the emission band in wavelength direction yields the error.

$$e_{\text{continuum}}(\rho) = \left(\frac{(\sum_i F_i^{\text{left}}) + (\sum_i F_i^{\text{right}})}{2} \right) \cdot d\lambda_{\text{Emission}} \quad (8.19)$$

Since both errors are independent the error of the measurement can be calculated from these two by the law of error propagation

$$e_{\text{Ges}} = \sqrt{e_{\text{total}}^2 + e_{\text{guiding}}^2} \quad (8.20)$$

Maximizing the signal-to-noise ratio

In nearly all nights several spectra of the comet have been obtained. For all observations except the last VLT/UT1 observations, the spectra have been reduced separately. The spatial profiles for the gas emission lines have been measured in each spectrum. To maximize the signal to noise ratio (S/N ratio) the mean of all spatial profiles obtained in the same night for a given radical has been computed. Different slit orientations have not been treated separately, because for this work a radial symmetric coma was assumed. The error of this mean profile was calculated from the errors on the single profiles using the law of error propagation.

In the case of very noisy data the combined profiles have been further processed by averaging the two spatial directions. Averaging the profiles was done on the assumption of radial isotropic outgassing from the nucleus. Both models used in this work, the Haser model and the ComChem model are based on the assumption of radial isotropic outgassing from the nucleus.

In general, the differences during the night and between sun- and tailward direction were small compared to the scatter within the spatial profile.

For the last observations at the VLT/UT1 ($r_h \geq 9.8$ AU) the spectra have been coadded after the correction for atmospheric extinction.

Converting observed fluxes to column densities

To convert the observed fluxes to column densities the conversion factor is defined as

$$f = \frac{s_{\text{proj}} \cdot p_{\text{proj}} \cdot g / r_h^2}{4 \cdot \pi \cdot \Delta \cdot \Delta}$$

s_{proj}	projected slit width on the sky [cm]
p_{proj}	projected pixel size on the nucleus [cm]
g	Fluorescence efficiencies (g-factor) at 1AU scaled with r_h in [AU] $[\frac{\text{erg}}{\text{s}^{-1}}]$ (see table B.1 and B.3 in Appendix B)
Δ	geocentric distance [cm]

The projected slit width and the projected pixel size, respectively, are obtained by simple geometric considerations.

The conversion from observed fluxes F in [10^{-16} erg cm $^{-2}$ s $^{-1}$] to column densities in [cm $^{-2}$] is finally done by

$$N = \frac{F}{f} \tag{8.21}$$

