PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC
101: 1032–1037, November 1989

THE EXTRACTION OF HIGHLY DISTORTED SPECTRA

T. R. MARSH

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218

Received 1989 June 13

ABSTRACT

Digital spectra of point sources can be extracted with maximum signal-to-noise ratio by applying appropriate weights across the profile of the object. The first stage in computing these weights involves determination of the fraction of the flux of the object which falls into each pixel. The published methods differ at this stage, and none can cope with some commonly available types of data. We present an improved method for fitting the object profile which is particularly suited to the extraction of tilted and distorted spectra, as produced by data taken with cross dispersers and image tubes.

Key words: data reduction–spectroscopy

1. Introduction

During the past two decades linear detectors of high quantum efficiency have been developed which enable accurate background subtraction even when the object is considerably fainter than the sky. Several authors have developed algorithms to extract such spectra with the optimum signal-to-noise ratio (Horne 1986; Robertson 1986; Urry and Reichert 1988). The same algorithms reduce the errors caused by poor flat-field correction and sky subtraction and can be used to identify cosmic-ray hits. In each case the first step is to compute the fraction of flux that falls into each pixel across the spectrum at every wavelength. In the second step these values are used to derive the optimum weights and, thus, to extract the spectrum.

The first step is the most difficult, and the published methods have all differed at this stage. Assuming that the spectra run parallel with the columns of a detector, Robertson (1986) used the fraction averaged over each column as the fraction at every wavelength. This fails if there is any tilt on the spectrum, for which Robertson suggested taking the average in several blocks in each of which the tilt was relatively small. However, this method leads to discontinuities between the blocks where the estimated fractions, and therefore the extraction weights, suddenly change. The more complex algorithm of Horne (1986) avoids this problem by fitting low-order polynomials along the columns. Besides coping with small tilts, the fitted profiles can then be used to detect bad pixels (e.g., cosmic-ray hits). For highly tilted spectra, which cross many columns, Horne's method fails because each column contains only a small slice of the spectrum and yet requires a high-order polynomial to fit it.

A third method was developed by Urry and Reichert (1988) to extract IUE spectra by fitting a Gaussian to several sections of the profile. The fitted widths and centroids were then smoothly interpolated for the entire spectrum. This method can cope with highly tilted spectra, but it is not clear that the profile will have a Gaussian shape, and, if not, accurate profile weights can never be computed.

In this paper I present an improved method which is the most direct extension of Horne's polynomial fitting method for tilted spectra. In the next section I summarize the optimal extraction method and discuss the current profile-fitting methods and the conditions under which they fail. I then describe the improved algorithm and I show its application to highly tilted data taken with a cross disperser and a CCD.

2. Optimal Extraction

We start by summarizing the equations which describe optimal extraction (following Horne 1986) and, hence, show the prime importance of fitting the fraction of flux that falls in every pixel. We assume that the image of the slit lies parallel to the rows of the detector and that the dispersion direction is roughly parallel to the columns. At a particular row of the detector let the amount of flux (with any background already subtracted) that falls into the ith pixel be D_i. If F is the true flux that we wish to estimate and P_i is the fraction which falls into pixel i, then $\Sigma_i P_i = 1$ and the expected value of D_i is given by $E(D_i) = P_i F$.

We estimate F by forming a weighted sum of the D_i across the profile, and so the estimate $\hat{F}$ is given by

$$\hat{F} = \sum_i W_i D_i \ , \qquad (1)$$

and, assuming that the weights are not correlated with the pixel fluxes, the expected value of our estimate is there-

fore

$$E(\hat{F}) = \sum_i W_i E(D_i) = F \sum_i W_i P_i \quad . \tag{2}$$

We impose the condition that our estimate is unbiased and therefore $\Sigma_i W_i P_i = 1$. Since $\Sigma_i P_i = 1$, the normal extraction method with unit weight for all pixels gives an unbiased estimate as expected; however, in general it is not the estimate of minimum variance. The variance of our estimate is given by

$$V(\hat{F}) = \sum_i W_i^2 V_i \quad , \tag{3}$$

where we denote by V_i the variance on pixel i. This expression can be minimized subject to the normalizing conditions using Lagrangian multipliers and we find

$$W_i = \frac{P_i/V_i}{\Sigma_i P_i^2/V_i} \quad . \tag{4}$$

As expected, pixels of larger uncertainty and containing a smaller fraction of flux are given the lowest weights. When the variance is dominated by the flux of the object then $V_i \propto D_i \propto P_i$, and the weights are uniform. Optimal extraction is useful if background noise is significant. Horne (1986) discusses the determination of the variance of each pixel which requires a model for the noise produced by the detector. For example, once flat-field variations are removed the noise produced by CCDs is well described by a readout noise component plus photon noise.

Assuming that the variances V_i are uncorrelated with the signal D_i, the weights derived above automatically satisfy the normalization condition for an unbiased estimate regardless of the variances. Therefore, although incorrect variances will degrade the signal-to-noise ratio, they do not bias the estimate. On the other hand, if the P_i are estimated incorrectly as P_i', say, then $\Sigma_i W_i P_i' = 1$, but in general $\Sigma_i W_i P_i \neq 1$ and therefore there will be some bias. This is only avoided if the weights are all equal, but since the optimum weights are not normally equal, correct estimation of the P_i is the most important step.

We now examine two of the current methods for fitting the profile. As remarked above, fitting an analytic function such as a Gaussian to the profile is a possible method, and it has been used with some success in the past (Urry and Reichert 1988). This method is simple to apply to tilted spectra since one needs only to know how the spectrum position perpendicular to the dispersion changes as a function of position along the spectrum. With this approach it is always possible that the function used is not a faithful representation of the true profile. This would occur if the spectrum had trailed along the slit, for example. An obvious extension would be to add in another function such as a Lorentzian profile. However, the more complex function may still fail, and one is faced with a series of ad hoc choices of functions which may need to be altered from spectrum to spectrum.

Horne's (1986) method removes this difficulty by fitting low-order polynomials along the columns and, therefore, makes no assumptions about the shape of the profile except that it varies smoothly with wavelength. This method is effective when the spectra are nearly parallel to the columns and when they do not suffer any short-scale displacements perpendicular to the dispersion. If these conditions hold, typically only three or four coefficients are needed for each column. While a particular profile shape is not assumed, this method does not extend as easily as fitting analytic functions to highly tilted spectra.

Consider the raw data frame displayed in Figure 1. This is cross-dispersed data which will be described in more detail in the next section. The spectra cross many columns, and a single column only covers a small section of each spectrum. It is especially difficult to fit data at the top end of the spectra where the signal is weak and only significant in a very few rows for any one column. In this region it is likely that the number of fit coefficients will be of the same order as the number of data points, which

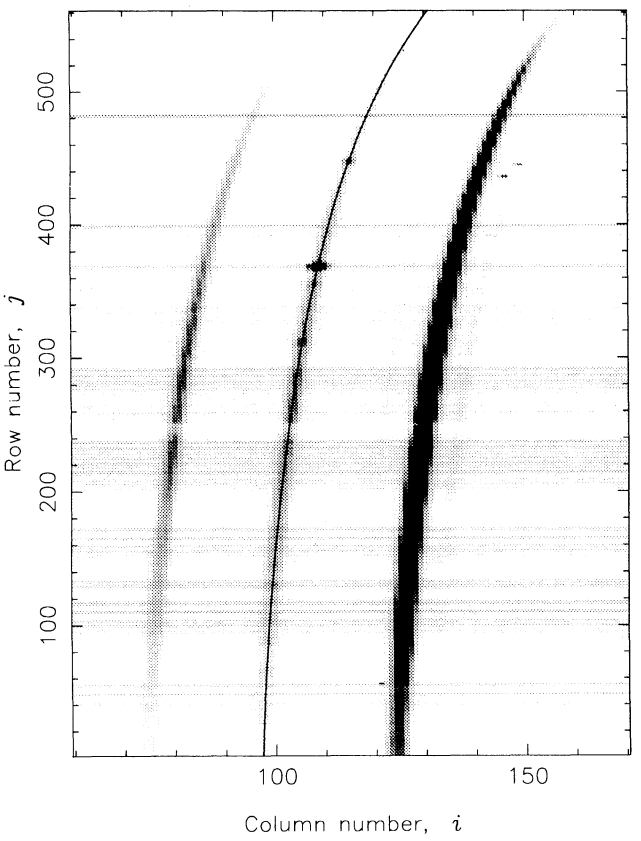

FIG. 1–A grey-scale image of three spectra with the FOS and WHT on La Palma. Only a section of the 400 × 590 frame is plotted. The middle star of the three is the eclipsing dwarf nova, HT Cas. A polynomial fit to the centroid of the spectrum has been plotted over it.

causes correlation between the data values and the estimated weights. The same problem occurs with image-tube spectra which, even if they include hardware to reduce distortion, typically show very short-scale distortion at either end of the spectrum.

We need a method which retains the advantages of Horne's method but also allows us to cope with tilted spectra. I present one such method in the next section, However, before proceeding I briefly consider an obvious technique that in fact cannot work.

If we can measure the tilt of the spectrum, then we could clearly remove it by resampling the data while applying the correct shifts at each wavelength. A spectrum could then be extracted from the straightened data frame. If the resampling algorithm conserves flux, then this method works for normal extraction. However, for optimal extraction it fails for two reasons. First, any resampling method leads to correlation between adjacent pixels and, therefore, we can no longer assume statistical independence during the extraction. Even without this effect, resampling always fails when the spectrum profile is undersampled. For example, consider a narrow profile which falls into a single pixel. In the absence of any other information, if such a profile were to be shifted by half a pixel, it must be divided equally between two pixels. However, although unavoidable, this is not necessarily the correct solution since the profile may not have been centered in the original pixel. The effect of this in practice is to cause short scale oscillations along the columns which are easy to miss in low signal-to-noise data but which cause periodic modulation of the extracted spectrum. I now consider a method which does not require resampling the data.

3. Fitting the Spectrum Profile

Variations with wavelength in the fraction of flux falling in a given column can come from true variations of the profile shape, as may be caused by a varying focus, or because of the changing position of the spectrum. For cross-dispersed and image-tube data, the latter is the most important, which suggests that since it is easy to fit the position of the spectrum we may be able to use the position to reduce the total number of fitting coefficients that we need.

This is the basis of the method. We start by fitting a polynomial to the position of the spectrum as a function of pixel number in the dispersion direction. A variety of methods are possible at this stage. We measure the position at each wavelength by cross correlation with a Gaussian (Schneider and Young 1980), a method which is less sensitive to noise than a simple centroid. The positions are then fitted with a polynomial on the assumption that the spectrum position varies smoothly. We will denote the spectrum position at row j by X_j, a smooth function of j.

Extending the notation of the previous section to include variation with wavelength or equivalently row number j, an initial, noisy estimate of the fraction of flux in the pixel in column i, row j is given by the signal in the pixel divided by the total flux in the row

$$E_{ij} = \frac{D_{ij}}{\Sigma_i D_{ij}} \quad . \tag{5}$$

We now imagine a set of polynomials which cover the object profile as in Horne's (1986) method with the difference that the polynomials are now evaluated on lines parallel to the fitted position of the spectrum rather than the columns. This means that the coefficients of one polynomial can be influenced by data in several columns and that an interpolation scheme is needed to translate the polynomial values into the object profile. Also in contrast to Horne's method, the same data may influence several polynomials, and so the fits to the polynomials cannot be made independently. Assume that there are K polynomials which are functions of the row number, are equally spaced along the rows, and are interpolated onto the correct position of the spectrum for any given row. The fitted fractions are then given by

$$P_{ij} = \sum_{k=1}^{K} Q_{kij} G_{kj} \quad , \tag{6}$$

where G_{kj} is the polynomial k at row j and Q_{kij} is an interpolation coefficient that determines the contribution of polynomial k to pixel (i,j). The coefficients are fixed by the position of the spectrum and by the interpolation method chosen. They are not altered by the fit. The fit is made by minimizing the χ^2 between the model and the data

$$\chi^2 = \sum_{ij} \frac{(E_{ij} - P_{ij})^2}{\sigma_{ij}^2} \quad , \tag{7}$$

where σ_{ij}^2 is the variance on the estimate E_{ij} and can be evaluated from equation (5) and the variances V_{ij} on each pixel. The sum is carried out over all pixels with significant object flux. Accounting for the spectrum position removes most of the profile variation, but in case of true shape changes, we still need to allow for some variation in the dispersion direction and so G_{kj} is given by a polynomial with N terms

$$G_{kj} = \sum_{n=1}^{N} A_{nk} j^{n-1} \quad , \tag{8}$$

where the NK coefficients A are the coefficients that we wish to fit. The advantage of this method is that N can remain small even for highly tilted spectra since most of the tilt is accounted for by the centroid fit included in the computation of Q_{kij} (see below). Solving for the coefficients A requires the minimization of equation (7), which is straightforward since the polynomials are linear in A, and we merely need to solve NK simultaneous equations

to find the A. The algebra demonstrating this is contained in the Appendix.

We now discuss the computation of the interpolation coefficients, Q_{kij}. The polynomials are intended to represent the profile before it is binned into pixels; to find the observed, binned profile, we must integrate the polynomials profile. The polynomials only give a finite number of samples of the unbinned profile and so some form of interpolation must be defined. Finally, when integrated, we can deduce the interpolation coefficients Q, which enter in equation (6). We now go through the scheme explicitly for the simplest interpolation method.

The polynomials are equally spaced along the rows parallel to the fitted position of the spectrum and so let polynomial k at row j be centered on

$$x_{kj} = X_j + Sk + C \quad , \tag{9}$$

where S is the spacing between the polynomials, X_j the position of the spectrum, and C an added constant to ensure that the polynomials cover the correct range across the profile. The spacing between the polynomials should be small enough to recover the unbinned profile. Therefore, it must be less than one to cope with undersampled data. We consider nearest-neighbor interpolation for the unbinned profile. In this case the unbinned profile is given by

$$P_{xj} = \sum_{k=1}^{K} \Pi(x - x_{kj}, S) G_{kj} \quad , \tag{10}$$

where the "top-hat" function $\Pi(x,S)$ is equal to 1 for $|x| < S/2$ and equal to 0 otherwise. The binned profile P_{ij} is the function P_{xj} integrated between $i - 0.5$ and $i + 0.5$. Carrying out this integration we find

$$P_{ij} = \sum_{k=1}^{K} \max\left(0, \min\left(S, \frac{S+1}{2} - |x_{kj} - i|\right)\right) G_{kj} \quad , \tag{11}$$

where $\max(a,b)$ returns the maximum value of its two arguments and $\min(a,b)$ returns the minimum. Comparing this expression with equation (6), we can identify the values of Q_{kij}. For this method of interpolation, the values of Q_{kij} for a given column i and polynomial k change continuously with X_j, but their derivatives with respect to X_j are not continuous, which causes gradient discontinuities in the fitted profile. Therefore, in practice we use linear interpolation of the unbinned profile according to

$$P_{xj} = \sum_{k=1}^{K} \Lambda(x - x_{kj}, S) G_{kj} \quad , \tag{12}$$

where $\Lambda(x,S)$ equals 0 for $|x| > S$, 1 at $x = 0$ and is linear between $x = 0$ and $x = \pm S$. The Q coefficients can be evaluated by integration as above.

For stability the first fit is made with equal variances on each pixel. In succeeding fits the previous fit is used to derive accurate variances, and so at least two fits are required. At the same time, deviant pixels can be detected by their deviation from the fit, and subsequently ignored. It normally takes five or six fits to eliminate all such pixels. This procedure is analogous to the outlier rejection method discussed by Horne (1986).

We now apply the method to data taken with a CCD and cross disperser.

4. Application to Data

The tilted profile fitting was designed for extraction of faint object spectrograph (FOS) data from the William Herschel telescope (WHT) on La Palma. We have also applied it to image-tube data taken with the 2-D Frutti on the 200″ (5-m) Hale telescope and to IPCS data taken on the Isaac Newton telescope on La Palma—each of which show short-scale distortion.

We now show the application to the FOS data. The FOS is a fixed-format instrument designed for high efficiency. Two orders separated by a cross disperser cover a wavelength range from 3500 Å to 9800 Å. We examine data taken during a run in July 1988. For these data the blue order was blocked with a filter, but since the cross-dispersing prisms are fixed in place, the red order (4900 Å to 9800 Å) remains strongly curved. The targets were cataclysmic variable stars which have orbital periods between 2 and 10 hours and, therefore, required short exposures to avoid smearing out orbital variations. Thus, the faintest parts of the spectrum are dominated by readout noise, and optimal extraction is useful.

Figure 1 shows the data which will be used to illustrate the profile fit. The middle star of the three on the slit is the eclipsing dwarf nova HT Cassiopeiae. The prominent emission lines of Hα at row 185 and He I 5876 at row 225 are clearly visible as are numerous night sky lines. The smooth line drawn through the data is the polynomial fit to the centroid of the spectrum at each wavelength. We used eight coefficients and the reduced $\chi^2 = 1.07$. The spectrum is tilted by 33 columns and low-order polynomial fits are inadequate. More details of the fits can be found in Table 1.

In Figure 2 we show the estimated fraction of flux falling into all rows of selected columns along with the fit based on linear interpolation plotted as solid lines over the top. We have not shown all the columns for clarity; the columns used are indicated by numbers which can be related to the X ordinate of Figure 1. Columns on the right-hand side cross the most tilted region of the spectrum and subsequently show more rapid variation than columns on the left. The maximum flux falls onto a given column when the spectrum is centered on it. In Figure 2 the maximum fraction in a column increases from left to right, which shows that the spectrum is best focused at the top of the image in Figure 1. This is the true profile variation which requires the number of polynomial terms N to be greater than 1.

Details of the fits can be found in the table. As can be

TABLE 1

Parameters of the Profile and Centroid Fits

Profile fit		Centroid fit	
Number of terms in each polynomial	3	Number of polynomial terms	8
Spacing, S	0.4 pixels	Reduced χ^2	1.07
Number of polynomials, K	34	RMS deviation	0.11 pixels
Rejection threshold	3σ		
Reduced χ^2	1.05		
Points rejected	87		
Total number of points	6905		

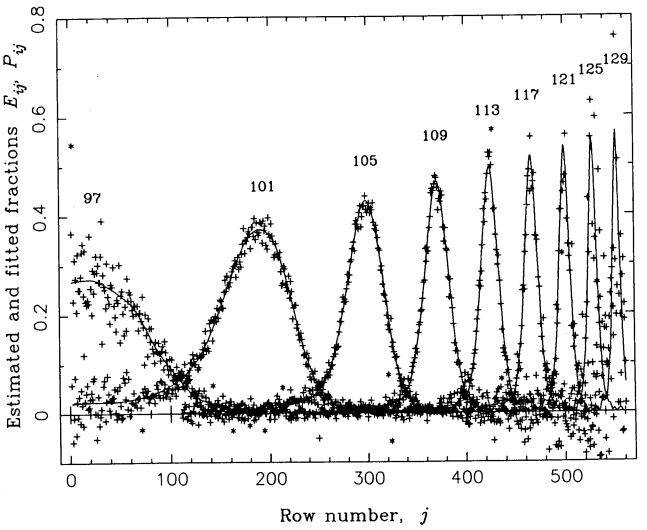

FIG. 2–The estimated fractions of flux falling in each pixel of selected columns are plotted along with the fit using the method of this paper. The stars are points rejected during the fit.

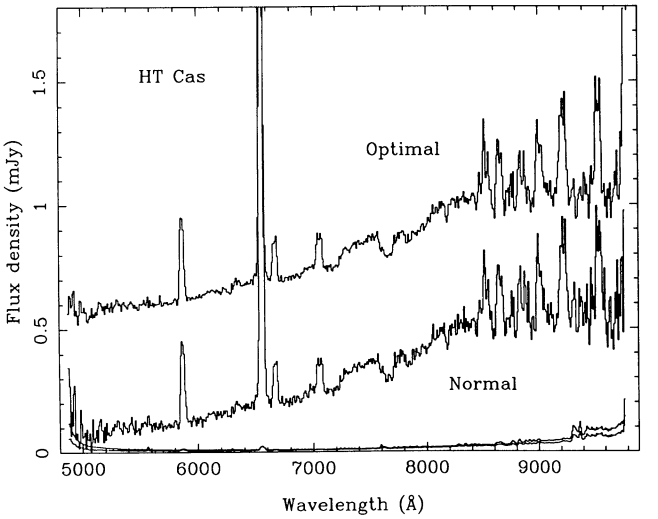

FIG. 3–The optimally and normally extracted spectrum of HT Cas from the frame of Figure 1 after applying flux and wavelength calibration and correcting for atmospheric absorption. Improvements to the signal-to-noise ratio are evident in many regions. The estimated 1 σ uncertainties are plotted along the bottom.

seen from Figure 2 and the reduced χ^2 in the table, the fit is excellent and provides a good estimate of the fractional flux in each pixel. The improvement over independent fits to each column is especially important for columns on the right-hand side, where the data are very noisy and could not be fitted without the information provided by other parts of the spectrum. The fitted profiles can now be used to extract the optimal spectrum, which requires only simple modifications to earlier algorithms. Figure 3 shows the optimal extraction based on the fit of Figure 2 compared to the normal extraction (unit weights) over the same region. The improvement in signal-to-noise ratio is evident in many parts of the spectrum.

5. Conclusion

Apart from the profile fitting, the new method is the same as Horne's (1986) method and, therefore, can be used to identify bad pixels on the spectrum in exactly the same way. In contrast to Horne's method, the more complex fitting presented here can allow the fit from one profile to be shifted onto another, since it only requires changing the fit to the spectrum position which is used in equation (9). This may be useful for optimal extractions of faint targets using profile fits from a bright target shifted to the new position.

In effect the fit is able to remove the effect of aliasing by using the many samples of the profile at different positions provided by the tilt. Recovery of spatial frequencies which give whole numbers of cycles within one pixel is not possible, of course, but this does not matter since it is the binned fit (which does not contain such frequencies) that is required for the optimal extraction.

I have written the algorithm as a set of VAX FORTRAN-77 subroutines, designed to run under the spectral reduction package FIGARO. Both the subroutines and the FIGARO programs which run them are available on request from the author.

I thank the Director of the Space Telescope Science

Institute for making funds available from the Director's Research Fund to support this work. I also thank Keith Horne for comments on earlier versions of this paper.

Appendix

In this Appendix I derive the explicit expressions needed to compute the coefficients A which minimize χ^2 in equation (7). We first differentiate equation (7) with respect to A_{nk} and set the result equal to 0:

$$\frac{\partial\chi^2}{\partial A_{nk}} = -2\sum_{ij}\frac{(E_{ij}-P_{ij})}{\sigma_{ij}^2}\frac{\partial P_{ij}}{\partial A_{nk}} = 0 \ , \qquad \text{(A1)}$$

and after substitution for P_{ij} from equation (6) and for the polynomials G from equation (10), and avoiding the repetition of symbols for unrelated indices, we find

$$\sum_{ml}A_{ml}\sum_{ij}\frac{Q_{kij}Q_{lij}j^{n+m-2}}{\sigma_{ij}^2} = \sum_{ij}\frac{E_{ij}Q_{kij}j^{n-1}}{\sigma_{ij}^2} \ . \qquad \text{(A2)}$$

The indices l and k range from 1 to K, the number of polynomials, and the indices m and n range from 1 to N, the number of coefficients for each polynomial. By defining two new indices, p and q, which uniquely specify all combinations of the four old indices, l, k, m, and n, we can transform the equation above into a standard matrix expression:

$$p = N(l-1)+m \ ,$$

$$q = N(k-1)+n \ ,$$

$$B_p = A_{ml} \ ,$$

$$X_q = \sum_{ij}\frac{E_{ij}Q_{kij}j^{n-1}}{\sigma_{ij}^2} \ ,$$

and

$$C_{qp} = \sum_{ij}\frac{Q_{kij}Q_{lij}j^{n+m-2}}{\sigma_{ij}^2} \ . \qquad \text{(A3)}$$

With these definitions equation A2 becomes

$$\sum_{p\ 1}^{NK}C_{qp}B_p = \mathrm{X_q} \ ,$$

for

$$q = 1,2,\ldots,NK \ , \qquad \text{(A4)}$$

which can be solved for the polynomial coefficients in B provided that the matrix C has an inverse. There are many algorithms available for the solution of simultaneous equations like equation (A4). Pixels can be ignored during the fit by leaving them out of the sum over i and j.

The computation of the matrix C takes the most time as it requires of order $(NK)^2 \times N_{\mathrm{P}}$ multiplications and additions where N_{P} is the number of pixels. This number can easily reach 10^8 although substantial savings are possible since many of the interpolation coefficients Q equal zero and since C is symmetric.

REFERENCES

Horne, K. 1986, *Pub. A.S.P.*, **98**, 609.
Robertson, J. G. 1986, *Pub. A.S.P.*, **98**, 1220.
Schneider, D. P., and Young, P. J. 1980, *Ap. J.*, **238**, 946.
Urry, M., and Reichert, G. 1988, *NASA IUE Newsletter*, **34**, 96.