

# Optimal Techniques in Spectroscopy: The Extraction of Multiple Spectra from Single and Multiple Exposures using Moments and B-Splines

Daniel D. Kelson

*The Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101*

kelson@ociw.edu

## ABSTRACT

Spectrograms are distorted in both the spatial and dispersion directions. As a result the data are not generally sampled on a uniform rectilinear grid of physically useful coordinates. Observers have often been required to resample their data onto rectilinear coordinate systems in order to continue with standard methods of analysis. The rebinning process involves replacing the data with an interpolating function, followed by sampling of that function at desired regular intervals of physically useful coordinates. These interpolating functions usually do not make full use of the information available in the data, and tend to degrade the resolution of spectra. With modern computing resources and knowledge of the distortions, one can construct interpolating functions that optimally reproduce the data. The information content of the data is preserved, without, for example, degradation of the resolution in the spectra. This paper discusses how to construct such interpolating functions for use in rebinning and extracting spectra. While the discussion is focused on the specific application for echellograms obtained with the MIKE spectrograph at Magellan, the method has been successfully applied to other instruments.

*Subject headings:* methods: data analysis — techniques: spectroscopic

## 1. Introduction

For more than a century key astrophysical questions have been tackled by making accurate measurements of physical quantities for astronomical objects. Numerous examples abound: measurements of galaxy redshifts have allowed us to deduce the presence of dark matter in clusters (Smith 1936, and subsequent work); measurements of stellar velocities

allow us to deduce the presence of planets around other stars (Marcy & Butler 1992, and subsequent work); measurements of weak absorption features of Uranium and other radioactive species allow us to deduce the ages of the oldest stars (Fowler & Hoyle 1960, and subsequent work); and measurements of the internal kinematics of galaxies allow us to deduce the masses and mass-to-light ratios of stellar systems (e.g., Burbidge et al. 1961a,b; Kelson et al. 2000, 2002, and others). Such conclusions may be startling, revolutionary, or mundane, but all such results depend critically on the statistical significance and accuracy of one’s measurements. Thus our beloved standard models, hypotheses, and theoretical underpinnings live and die by our ability to estimate uncertainties and make optimal use of the available data. At every step in the long process of converting a set of observations to reduced quantities, uncertainties are always added, information content tends to be reduced; entropy increases. It behooves the astronomer to both minimize and estimate the added uncertainties, though sometimes it is easier to do one than the other, and to take care not to weaken the value of the observations through excessive degradations.

The examples given above, though by no means exhaustive, illustrate key areas of astrophysics that depend on the careful analysis of spectrograms. One of the last steps in such analysis, before one can obtain reduced astrophysical quantities, involves extracting the spectrum of an object from a two-dimensional data array. The task of extracting one-dimensional spectra has become routine, with some variation among the commonly used procedures (e.g. Horne 1986; Robertson 1986; Marsh 1989; Mukai 1990). Some procedures either assume there is alignment of one of the principal axes of the data with the columns or rows of the detector or altogether require “rectification” before (or after) modeling and subtracting a background spectrum from the data (see, e.g. Kelson 2003, for a discussion of background-subtraction). Often there are large distortions in the data, and line-curvature can be sufficiently large to reduce the effective resolution of extracted spectra when a rectification step is ignored. It is not uncommon for line-curvature to be ignored when it is only a modest effect but this behavior is not desirable as it degrades the observations by worsening the effective resolution of the extracted spectra. Such procedures do not make full use of the information contained in the data and thus could hardly be called “optimal.”

Subsequent analyses of extracted spectra often require a sampling at regular wavelength intervals, thus necessitating the resampling of either the two-dimensional spectrum prior to extraction or the rebinning of the extracted one-dimensional spectrum.<sup>1</sup> This paper will discuss a new method for extracting one-dimensional spectra with the aim of producing spectra

---

<sup>1</sup>While some analyses do not require spectra to be sampled at regular wavelength intervals, it has become commonplace to work with spectra rebinned, for example, logarithmically in wavelength for measurements of radial velocities or for measurements of internal kinematics.

that may be sampled on a regular coordinate system but, more importantly, with minimal loss of information. The algorithm involves fitting for parameterizations of the data within the (original) distorted coordinate system. The fitting is performed using standard linear least-squares routines, with weighting using the expected noise in the data. By minimizing the degradation of information within the data, and by weighting the pixels appropriately, we call the method both “maximal” and “optimal.”

Rebinning either two-dimensional or one-dimensional spectra is often thought of in simple terms. In actuality this task is a complicated convolution, one that is sufficiently complex to be generally irreversible. The process is destructive; the observations are *replaced* by an interpolating function and that function is subsequently sampled at coordinate intervals that are typically strongly non-linear functions of the intrinsic coordinate system of the detector. Usually the desired coordinate system has regular wavelength intervals, and in the case of two-dimensional spectra the data are usually resampled onto a regular spatial coordinate system as well. The new two-dimensional rebinned spectra have pixel values that come from the interpolating function with no guarantee that the interpolating function optimally represents the data. The remainder of the data reduction and subsequent analysis thus become strongly dependent on the quality and characteristics of the interpolating function. For data that are not well-sampled (spatially or spectrally), the interpolating function may not even have mathematically continuous first derivatives. The noise characteristics of such data can become quite complicated, requiring tremendous overhead to account for the modifications to the variances (e.g. Cardiel et al. 1998; Kelson et al. 2001, 2005). Worse still, ugly artifacts can be introduced into the data. By the end of a pipeline, the resulting spectra may have very different structural characteristics on scales of a pixel compared to the underlying spectrum that was observed.

Increased computing capacity now allows us to devise more sensible interpolating functions. This paper introduces a method for constructing optimal interpolating functions for the extraction and resampling of spectra. We focus the discussion on unresolved and marginally-resolved continuum sources, though the procedures can be modified for extended sources. More specifically, the context for our introduction of the algorithm is our implementation for echellograms taken with the MIKE spectrograph (Bernstein et al. 2003) on the Clay 6.5m at Magellan. However, the procedures are straightforward to implement for any echelle spectrograph, and can easily be adapted for single-slit single-order spectroscopy or for multi-slit spectroscopy, substituting dependencies on aperture plate coordinates for dependencies on spectral order.

At its most basic level the interpolating function is no longer a function of CCD coordinates (i.e. row and column) but is constructed as a separable function of “rectified”

coordinates (i.e. unique spatial and wavelength-dependent coordinates for each pixel). Efficient linear algebra libraries make the computations sufficiently easy so as to prevent the reduction from being overly burdensome for modern computers with at least two gigabytes of memory. We now give an overview of the procedure.

## 2. The Method

Two-dimensional background-subtracted spectra suffer from two problems that must be dealt with before or during the process of extracting spectra: (1) the fact that the spectra are not aligned exactly along the rows (or columns) of the detector and are often curved with respect to the natural coordinate system of the detector (the y-distortion); and (2) the general tendency for dispersers to impose wavelength-dependent line curvature onto the two-dimensional spectra (which may have already been tilted or curved if the slit has been so cut into the aperture plate). Furthermore, coarse pixels have typically limited the accuracy with which one could previously deal with (1) and (2).

As in Kelson (2003), we first define the image of two-dimensional spectroscopic data as  $R(x, y)$ , where  $(x, y)$  are detector coordinates in pixels. Because of distortions imposed by the optics, the spatial coordinate along the slit,  $y_t$ , for a given pixel  $(x, y)$  is a non-linear function,  $y_t = Y(x, y)$ . Furthermore, the wavelength of light,  $\lambda$ , incident onto a pixel  $(x, y)$  is also a non-linear function of image position. For the purposes of constructing a sensible interpolating function for  $R(x, y)$  (i.e. modeling the data), we are less concerned with the actual wavelength,  $\lambda$ , of incident light than we are with the fact that there exists a coordinate system,  $(x_r, y_t)$ , in which  $x_r$  is a wavelength-dependent coordinate that is orthogonal to the spatial coordinate  $y_t$ . The transformation to this system is  $x_r = X(x, y_t)$  such that the wavelength of light incident on a pixel can be written  $\lambda = L(x_r)$ , where  $x_r = X(x, Y(x, y))$  and  $0 \leq x_r \leq \text{NAXIS1}$  (for data arranged such that the dispersion is aligned along the rows of the array). Thus there exists a convenient coordinate system  $(x_r, y_t)$  for which  $\partial L / \partial y_t = 0$ . For echellograms  $Y$  is generally referred to as the order curvature and it can be measured accurately from the edges of quartz lamp or twilight spectra.

Figure 1 illustrates the  $(y_t, x_r)$  coordinate systems. In the top panels we show,  $y_t = Y(x, y)$ , the order curvature of a MIKE exposure. The bottom panels show a subsection of the same MIKE data along with contours in both  $y_t$  and  $x_r$ . The mapping of the line curvature,  $X(x, y_t)$ , can be accurately measured using comparison lamp spectra or night-sky emission features (e.g. Kelson et al. 2000). Note that while  $x_r$  and  $y_t$  are physically orthogonal coordinates, contours of  $x_r$  and  $y_t$  are typically not orthogonal in the  $(x, y)$  plane. Other, non-instrumental, effects may also affect the relationship between these coordinate systems

and coordinates on the sky (later to be referred to as  $y_c$ ). Some of these will be discussed below but for now we only define the instrumental coordinate transformations.

We assume that before any such extraction procedures are performed, the data have been bias-subtracted, flat-fielded, and background-subtracted (e.g. Kelson 2003).

## 2.1. Localization of Spectra

The first step is to accurately localize the object(s) to be extracted. Since we are introducing the method in the context of echelle spectroscopy, we present a simple technique for accurately localizing the position of the object in all orders simultaneously (similar techniques can be used in multi-object spectroscopy as well). Because of effects such as atmospheric dispersion, object spectra may not actually follow the order curvature as measured by  $Y$  in the previous paragraph. The procedure discussed here for localizing the object spectra will fully take such deviations from the order curvature into account.

Within a single order (or slitlet) the mean (spatial) location of an object can be calculated by finding the first moment of the intensity distribution:

$$\langle y_t \rangle = \frac{\sum_{x_r, y_t} y_t R(x_r, y_t)}{\sum_{x_r, y_t} R(x_r, y_t)} \quad (1)$$

The first moment of an object with modest signal-to-noise ratio may be strongly affected by cosmic-rays and other non-Gaussian sources of bad pixel values. In these cases it is often preferred to use  $R_M(x_r, y_t)$ , a median-filtered version of  $R(x_r, y_t)$  using a kernel substantially elongated along  $x$ , the wavelength-dominated coordinate in the detector array.

We note that computing the first moment of the object is mathematically identical to a least-squares fit for the mean  $y_t$ , in which one using  $W_M(x_r, y_t)$  for weights:

$$\chi^2 = \sum_{x_r, y_t} [W_M(x_r, y_t)(y_t - A)]^2 \quad (2)$$

$$W_M(x_r, y_t) = \sqrt{R_M(x_r, y_t)} \quad (3)$$

This is the simplest model for the trace of an object, in which the first moment is simply a constant:  $A = (\sum W_M^2 y_t) / \sum W_M^2 \equiv \langle y_t \rangle$ .

That formal identity between  $A$  and the first moment of the intensity distribution (more specifically the distribution of  $W_M$ ) forms the basis of our method for localization. Ideally, the objects should trace lines of constant  $y_t$  in the data, However atmospheric dispersion and

errors in the fitting of  $Y$  can lead to deviations from the ideal. Thus, we must use a higher-order description for an object’s trace. We choose to represent the trace with higher-order polynomials. For example, the next simplest approximation for the deviation of the object’s “trace” from that of being parallel to the fit for the order curvature would be to compute the first moment  $\langle y_t \rangle$  as a linear function of  $x_r$ . For reasons that will become apparent later, we transform the wavelength-dependent coordinate such that  $x_l = (x_r - x_c)/x_c$  where  $x_c = \text{NAXIS1}/2$ . Therefore,  $-1 \leq x_l \leq 1$ .

One can solve for the dependence of the first moment as a linear function of the rescaled wavelength-dependent coordinate by minimizing

$$\chi^2 = \sum_{x_r, y_t} \{W_M(x_r, y_t)[y_t - (A + Bx_l)]\}^2 \quad (4)$$

and solving for  $A$  and  $B$ :

$$B = \frac{\langle x_l y_t \rangle - \langle x_l \rangle \langle y_t \rangle}{\langle x_l^2 \rangle - \langle x_l \rangle^2} \quad (5)$$

$$A = \langle y_t \rangle - B \langle x_l \rangle \quad (6)$$

For an object with flux symmetrically distributed about  $x_l = 0$ , one obtains  $A = \langle y_t \rangle$  and  $B = \langle x_l y_t \rangle / \langle x_l^2 \rangle$ .

The computational machinery for performing linear least-squares can be used to calculate the first moment of the intensity distribution in arbitrary coordinate systems. In our case we can calculate the trace(s) of the object spectra within the  $y_t$  coordinate system. We do so by parameterizing the first moment of the intensity distribution as a global function of independent coordinates. For echellograms we prefer to parameterize the first moment of the flux distribution as a two-dimensional polynomial of wavelength and spectrum order,  $a$ )<sup>2</sup>. For multi-object spectroscopy, using either fibers or slitlets, one can use the positions in an aperture plate as the independent coordinates on which to base the dependencies of the first moments.

We choose to fit Legendre polynomials for the first moment. We defined  $x_l$  above and now define the rescaled spectrum order  $a_l$ , where  $a_l = (a_o - a_c)/a_w$ ,  $a_c = \langle a \rangle$  and  $a_w = [\max(a) - \min(a)]/2$  such that the parameterization for the first moment is found by minimizing:

$$\chi^2 = \sum_{x_r, y_t} \left\{ W_M(x_r, y_t) \left[ y_t - \sum_{i=0}^I \sum_{j=0}^J \beta_{i,j} L_i(a_l) L_j(x_l) \right] \right\}^2 \quad (7)$$

---

<sup>2</sup>For typical MIKE blue and red spectra, the data span  $70 \lesssim a \lesssim 100$  and  $40 \lesssim a \lesssim 70$ , respectively.

where we typically use  $I = 6$  and  $J = 6$  for MIKE observations.

The computation of the trace of the first moment, as a function of spectrum order, allows us to define a spatial coordinate referenced to the center of the object at every wavelength and in every order:

$$y_c = y_t - \langle y_t \rangle \quad (8)$$

$$\langle y_t \rangle = \sum_{i=0}^I \sum_{j=0}^J \beta_{i,j} L_i(a_l) L_j(x_l) \quad (9)$$

Figure 2 shows a subsection of MIKE data, with dotted lines indicating contours of constant  $y_t$ . The solid line indicates the location of the first moment,  $\langle y_t \rangle$ . Note the curvature of the first moment with respect to the contours of  $y_t$ . In this way, errors in the curvature mapping, atmospheric dispersion, and other sources of deviation from the fitted order curvature are properly accounted for.

## 2.2. Parameterization of Spatial Profiles

Parameterization of the spatial profile, as a complicated function of  $x_l$  and  $a_l$  has several steps, the first of which is to empirically calculate the width or scale of the object. Fortunately, we do not need to perform a non-linear least-squares fit of, e.g., a Gaussian or Lorentzian to the spatial profile. We find that a Gauss-Hermite decomposition of the spatial profile works very well and is efficient to perform.

The first step in the Gauss-Hermite decomposition is to define the Gaussian that best matches the data. Normally the fitting of a Gaussian to an intensity distribution is performed using non-linear least-squares methods, such as commonly done using a modified Levenberg-Marquardt method (e.g., Press et al. 1992). However, a more efficient method is to apply the machinery of the previous section and parameterize the second moments of the flux distribution as a polynomial function of the order and wavelength:

$$\chi^2 = \sum_{x_r, y_t} \{ W_M(x_r, y_t) [y_c^2 - \sum_{u=0}^U \sum_{v=0}^V \gamma_{u,v} L_u(a_l) L_v(x_l)] \}^2 \quad (10)$$

We typically use  $U = 2$  and  $V = 2$ .

The values of  $\beta_{i,j}$  and  $\gamma_{u,v}$  determined at this point are not yet accurate enough to proceed. However a second pass through the computation of the first and second moments, restricting the calculation to those pixels with  $|y_c| < 2\langle y_c^2 \rangle^{1/2}$ , does result in sufficiently

robust determinations of  $\beta_{i,j}$  and  $\gamma_{u,v}$ . The restriction to those pixels with  $|y_c| < 2\langle y_c^2 \rangle^{1/2}$  is only used in the second calculation of the  $\beta_{i,j}$  and  $\gamma_{u,v}$ . One might naively think that a cut on pixels inside twice the estimated second moment is overly strict. However the first pass tends to over-estimate the second moment by about a third or so. Nevertheless, the remainder of the algorithm does not require this restriction.

With the second moment of the object parameterized as a function of spectrum order, one can now construct rescaled spatial coordinates

$$y_s = y_c / \langle y_c^2 \rangle^{1/2} \quad (11)$$

$$\langle y_c^2 \rangle = \sum_{u=0}^U \sum_{v=0}^V \gamma_{u,v} L_u(a_l) L_v(x_l) \quad (12)$$

to begin the decomposition of the spatial profile itself. Note that  $\langle y_c^2 \rangle$  is currently a function of spectrum order and wavelength. In Figure 2 we show the second moment using dashed lines. For multi-slit spectroscopy one would parameterize the second moment as functions of coordinates in the aperture plate.

For unresolved and marginally-resolved sources we also prefer to decompose the background-subtracted data,  $R$ , into two separate components (see also, e.g., Mukai 1990):

$$R(x_r, y_t) = S(x_r, y_s, a_l) P(y_s, a_l, x_l) \quad (13)$$

where  $S(x_r, y_s, a_l)$  is the two-dimensional spectrum of the object and  $P(y_s, a_l, x_l)$  is the spatial profile of the object. Each order has its own spatial profile and because we include a polynomial expansion of  $x_l$ , the spatial profile is allowed to vary with wavelength within each order. Unresolved sources allow us to simplify Eq. 13 to

$$R(x_r, y_t) = S(x_r, a_l) P(y_s, a_l, x_l) \quad (14)$$

These equations define the interpolating function we are attempting to construct. Everything up until this point in the analysis has been necessary for calculating  $y_c$  and  $y_s$ .

Before solving for  $S$ , we must first determine  $P$ . In our experience with data obtained using MIKE, LRIS, LDSS2, LDSS3, DEIMOS, and IMACS, unresolved and marginally-resolved sources have spatial profiles well-described by a finite Gauss-Hermite series:

$$P(y_s, a_l, x_l) = \sum_{e=0}^E \sum_{f=0}^F \sum_{g=0}^G \epsilon_{e,f,g} L_e(a_l) L_f(x_l) H_g(y_s) e^{-y_s^2/2} \quad (15)$$

where  $G$  is typically set to 10 but can be restricted such that  $P$  is a pure Gaussian ( $K = 0$ ). For our MIKE echellograms we choose values of 1 or 2 for  $E$  and 1 or 2 for  $F$ . As can

be seen above, we have parameterized the Gauss-Hermite moments of the spatial profile as Legendre polynomials of both  $a_l$  and  $x_l$ . For multi-object spectroscopy, one implements a dependence on metric coordinates in an aperture plate instead of spectral order. Note that we have constructed a parameterization in which the Gauss-Hermite moments of the spatial profiles of the object spectra are themselves dependent on order and wavelength (i.e. the wavelength-dependent coordinate). Thus we ensure that the spatial profile can properly take into account changes in focus across the detector.

In order to solve for  $\epsilon_{e,f,g}$  we first adopt the approximation

$$P'(y_s, a_l, x_l) = e^{-y_s^2/2} \quad (16)$$

and use Dierckx (1993) to find the B-spline  $S'(x_r, a)$  that minimizes

$$\chi^2 = \sum_{x_r, y_t} \left[ \frac{R(x_r, y_t)/P'(y_s, a_l, x_l) - S'(x_r, a)}{\tau(x_r, y_t)} \right]^2 \quad (17)$$

$$\tau^2(x_r, y_t) = R(x_r, y_t)/[\alpha P'(y_s, a_l, x_l)] + B(x_r, y_t)/[\alpha P'(y_s, a_l, x_l)^2] + \{\rho/[\alpha P'(y_s, a_l, x_l)]\}^2 \quad (18)$$

where  $\alpha$  converts counts to electrons ( $e^-$ ),  $B$  is the two-dimensional background image (computed in the rectified coordinate system, *a la* Kelson 2003), and  $\rho$  is the read noise (in  $e^-$ ). If the distortions are sufficiently small, then each order (or slitlet) may fully extend over the same range of  $x_r$  and thus  $S'$  can be fit as a bivariate function of  $(x_r, a)$ , with a knot spacing of 1 pixel in  $x_r$ , and an order knot spacing of 1 in  $a$ . However, some instruments may have sufficiently large order curvature for a few orders to have significantly less coverage in  $x_r$ . In these cases one must fit univariate B-splines in each order separately to construct  $S'$ . MIKE is one such instrument that requires the order-by-order calculation of  $S'$ . With *a priori* knowledge of the blaze function, one could incorporate it and solve for a single, order-independent  $S'$  as a function of wavelength, instead of  $x_r$ .

The approximation  $S'$  can now be used to calculate the Gauss-Hermite moments,  $\epsilon_{e,f,g}$ , by minimizing

$$\chi^2 = \sum_{x_r, y_t} \left[ \frac{R(x_r, y_t) - \sum_{e=0}^E \sum_{f=0}^F \sum_{g=0}^G \epsilon_{e,f,g} L_e(a_l) L_f(x_l) H_g(y_s) e^{-y_s^2/2} S'(x_r, a)}{\sigma(x_r, y_t)} \right]^2 \quad (19)$$

The resulting  $\epsilon_{e,f,g}$  coefficients for the polynomial representation of the Gauss-Hermite moments of the spatial profile. If the first and second moments have been computed with perfect accuracy then  $\epsilon_{e,f,1}$  and  $\epsilon_{e,f,2}$  will be identically zero. In practice, however, the polynomial representations for the first and second moments will not be perfect and these coefficients will not be perfectly equal to zero.

### 2.3. Calculation of Spectra

With the coefficients  $\epsilon_{e,f,g}$  now computed, the spatial profile of the object, in every order and at every wavelength, is given by Equation 15. Thus, we now proceed order-by-order (or slitlet-by-slitlet), and fit B-splines to  $S = R/P$  (Dierckx 1993). For unresolved sources one can fit a univariate B-spline, using only  $x_r$  as the independent variable. For marginally-resolved sources, or simply to compensate for uncertainties in the spatial profile, one can define  $S$  as a bivariate B-spline, found by minimizing

$$\chi^2 = \sum_{x_r, y_t} \left[ \frac{R(x_r, y_t)/P(y_s, a_l, x_l) - S(x_r, y_s, a)}{\tau(x_r, y_t)} \right]^2 \quad (20)$$

$$\tau(x_r, y_t)^2 = R(x_r, y_t)/[\alpha P(y_s, a_l, x_l)] + B(x_r, y_t)/[\alpha P(y_s, a_l, x_l)^2] + \{\rho/[\alpha P(y_s, a_l, x_l)]\}^2 \quad (21)$$

With the fitting for  $S$ , we now have an interpolating function,  $S(x_r, y_s, a)P(y_s, a_l, x_l)$ , that mimics the data. It has been constructed in a physically motivated coordinate system — wavelength and location along a slit — instead of the coordinate system of the detector. By doing so, we have fully exploited the variation in the sampling, due to line curvature and other distortions, of the spectrum  $S$ . Thus  $S$  has fully preserved the intrinsic resolution of the instrument, without the degradation caused by simple interpolation schemes.

Using this interpolating function, it is trivial to construct the equivalent of a one-dimensional “extracted” spectrum. One integrates  $SP$  over  $y_c$  at the desired wavelength intervals. Because  $S$  has been defined along lines of constant  $x_r$ , our resulting interpolating function (or model for the data) contains all of the resolution of the spectrum intrinsic to the data. Resampling this model does not degrade the resolution of the spectrum, unlike simple traditional interpolation functions.

If one is only interested in the extracted one-dimensional spectra, then the integral of  $SP$  is fairly insensitive to the choice of fitting a univariate or bivariate B-spline for  $S$ . The bivariate B-spline does serve the purpose of fitting out deviations from the average spatial profile, and for correcting other errors in the fitting of the first and second moments. However, if the trace computed in §2.1 is reasonably accurate, then the mean of  $S$  over the aperture of integration is mathematically conserved (though this is only strictly true if the weighting in Equation 21 does not have a large gradient in  $y_c$ ).

The adoption of a bivariate B-spline for  $S$  is more computationally expensive but it does also allow the effective spatial profile of the data to vary from the global prescription. In such cases, one is fitting for deviations from the average spatial profile at every wavelength. This enables, for example, extended emission lines to have spatial profiles that are substantially

different from the overall continuum. The fitting of  $S$  as a bivariate B-spline does take some additional care over the univariate case. One must ensure that the spatial extent of pixels included in the fit spans more than  $(2k_y + 2)$  points (or more, if interior knots are selected) where  $k_y$  is the order of the B-spline in the spatial direction.

For extended sources the Gauss-Hermite decomposition and bivariate B-spline fit for  $S$  may not be appropriate. One can substitute other forms such as de Vaucouleurs profiles, exponentials, univariate B-splines, or almost anything for  $P$ , so long as the deviations from that profile at a given wavelength are not too large. In such cases, the data could be fit using a quintic B-spline or one can employ more complicated tactics that involve the placement of interior knots along the spatial dimension (see Dierckx 1993, for discussions of various strategies for knot placement). We have only begun initial experiments on extended sources and so far these have been encouraging.

Performing the extractions by “forward” modeling the data allows one to explicitly test whether one’s interpolating function is an optimal representation of the data by explicitly minimizing  $\chi^2$ . However, the interpolating function can be sampled at arbitrary coordinates and thus one can visually inspect the quality of the fit/extraction. In Figure 3 and 4 we show the results of a fit to MIKE data for a bright target. In this case we used a univariate B-spline for  $S(x_r, a)$ . The top panels show the data, and the middle panels show the interpolating function sampled onto the identical grid as the data. The bottom panels show the residuals from the fit. In Figures 5 and 6 we show the fit to a faint source, in which cosmic-rays are present in significant numbers (though iteratively down-weighted in the fitting for  $S$ ).

## 2.4. Extension to Multiple Integrations

The decomposition of spectra using a spatial profile allows one to expand the fit for  $S$  to include pixels from multiple integrations. One fits for  $S$  by minimizing the residuals from  $M$  exposures:

$$\chi^2 = \sum_{m=1}^M \sum_{x_r, y_t} \left[ \frac{R_m(x_r, y_t) / [p_m P_m(y_s, a_l, x_l)] - S(x_r, y_s, a_l)}{\tau_m(x_r, y_t)} \right]^2 \quad (22)$$

$$\tau_m(x_r, y_t)^2 = \frac{R_m(x_r, y_t) / [\alpha p_m P_m(y_s, a_l, x_l)] + B_m(x_r, y_t) / [\alpha p_m P_m(y_s, a_l, x_l)]^2}{\{\rho / [\alpha p_m P_m(y_s, a_l, x_l)]\}^2} \quad (23)$$

The values of  $R_m/P_m$  from multiple exposures need to be appropriately scaled to account for variations in exposure time and atmospheric transparency. The scale factors,  $p_m$ , can be

computed, for example, by using the median values of  $S'$  in each exposure.<sup>3</sup> This strategy has been very useful in our reductions of MIKE, LDSS3, DEIMOS, and IMACS data because it facilitates the rejection of cosmic-rays and it makes maximal use of the available information regarding the small exposure-to-exposure variations in wavelength sampling.<sup>4</sup>

To illustrate the power of the multi-exposure fitting, we actually used this method in constructing Figures 3 through 6. The bright source in Figures 3 and 4 was fit using three exposures. The faint source in Figures 5 and 6 was fit using six. In this way, cosmic-rays are easily identified, down-weighted, and excluded from the fit for  $S$ . *Cosmic-rays never need to be replaced; they can be identified and merely ignored in the construction of the interpolating function.*

## 2.5. Summary of Method

We have constructed an interpolating function using global parameterizations for the first and second moments of the intensity distributions within each order of a spectrum, global parameterizations for the Gauss-Hermite moments of the spatial profile, and B-splines for the spectra within each order (or slitlet). These calculations are performed in each available exposure to normalize the counts for the final calculation of a B-spline representation for spectra that fit all exposures simultaneously. Fitting the traces and spatial profiles of all orders in an exposure with a single, global parameterization provides tremendous power for recovering spectra when the flux is far from a uniform distribution. Figure 7 shows this power using the results from fitting four exposures of a  $z = 5.78$  QSO. The figure shows a region with wavelengths ranging from 7200Å to 8200Å. There is very little flux at these wavelengths, but the flux is distributed over a broad (effectively random) range of  $x_r$  over the span of available orders. As a result the global parameterization for the trace and spatial profile is well-constrained. The final panel in the figure shows that no systematic residuals in the fit remain.

In our method, we have replaced the data with a much more complicated interpolating function than has normally been used in standard interpolation methods and we expect that it can accurately reproduce the data (without degrading the spatial and spectral resolution). Our tests with high signal-to-noise ratio stellar spectra, as illustrated with the previous figures, bear this out.

---

<sup>3</sup>We prefer to define  $p_m$  order-by-order (using  $S'$ ) to account for mild wavelength-dependencies.

<sup>4</sup>Note that in order to combine multiple exposures, one must transform the  $x_r$  coordinates of the various frames to a consistent wavelength system.

### 3. Summary

We have presented an algorithm for constructing spectroscopic interpolating functions that preserve the information content of observations. We use the machinery of least-squares minimization to perform several moment calculations for localizing all spectra within a data array, and for globally determining the spatial profiles of all of those spectra. Once the spatial profiles for all spectra are known, we use bivariate B-splines to fit the residual spectra. The resulting models of the observations are accurate representations in a least-squares sense, making full use of the information available over the full field of the data.

The global fit for the parameters that describe the spatial profiles in each order (or slitlet, in the case of multi-slit data) should be an important component of any method for extracting spectra from echellograms or from multi-object data. Orders or objects which contain low flux levels can still provide optimally extracted spectra now that their spatial profiles can be inferred from the rest of the data. For multi-slit data, the parameterizations can be expressed as functions of the coordinates in the multi-slit aperture plate, and objects too faint to provide an accurate individual determination of its own spatial profile can still be “optimally” extracted using a spatial profile determined from all objects obtained within an exposure.

Finally, the framework presented in this paper can be used as the basis for more complicated spectral extractions: that of extracting a single spectrum from multiple exposures. Individual exposures should have their own global parameterizations for their spatial profiles of the spectra. Then a single B-spline calculation can be made to find the average spectrum from all observations of a target. No rebinning of any data to a common coordinate system is required and the fit for the average spectrum can make full use all of the data in one shot. The small shifts between exposures, both in wavelength and spatially, provide additional constraints on the structure of the spectra on scales smaller than a pixel. The fit for the spectrum can be weighted according to the expected noise, and individual exposures can be weighted according to the observed counts or exposure times. By taking advantage of the information present in all of the data, one can finally construct interpolating functions that optimally match the observations, and by iterating the fit, one’s final interpolating functions can be free from the cosmic-rays and other types of bogus data. With the computational machinery available today, one no longer is required to rebin individual exposures and extracted spectra before combining the results. With proper and careful parameterizations of the spatial profiles of one’s spectra, found by fitting all spectra within each exposure simultaneously, one can avoid any degradation of one’s spectroscopic data.

The author would like to thank the Carnegie Observatories for supporting innovation in

science and research. Appreciation is also expressed for the patience of Ian Thompson, Andy McWilliam, Luis Ho, Jon Fulbright, and others in the long process of refining the method. Further thanks to Steve Sheckman, Ian Thompson, Andy McWilliam, Marla Geha, and Mike Blanton for suggestions that greatly improved the manuscript and its presentation.

## REFERENCES

- Beers, T. C., Flynn, K., & Gebhardt, K. 1990, *AJ*, 100, 32
- Bernstein, R., Sheckman, S. A., Gunnels, S. M., Mochnacki, S., & Athey, A. E. 2003, *Proc. SPIE*, 4841, 1694
- Burbidge, E. M., Burbidge, G. R., & Fish, R. A. 1961, *ApJ*, 133, 393
- Burbidge, E. M., Burbidge, G. R., & Fish, R. A. 1961, *ApJ*, 134, 251
- Cardiel, N., Gorgas, J., Cenarro, J., & Gonzalez, J. J. 1998, *A&AS*, 127, 597
- Cushing, M. C., Vacca, W. D., & Rayner, J. T. 2004, *PASP*, 116, 362
- Dierckx, P., *Curve and Surface Fitting with Splines*, New York: Oxford University Press, 1993
- Horne, K. 1986, *PASP*, 98, 609
- Fowler, W.A., & Hoyle, F. 1960, *Ann. Phys.*, 10, 280
- Kelson, D.D., Illingworth, G.D., van Dokkum, P.G., & Franx, M. 2000, *ApJ*, 531, 159
- Kelson, D. D., Illingworth, G. D., Franx, M., & van Dokkum, P. G. 2001, *ApJ*, 552, L17
- Kelson, D. D., Zabludoff, A. I., Williams, K. A., Trager, S. C., Mulchaey, J. S., & Bolte, M. 2002, *ApJ*, 576, 720
- Kelson, D. D. 2003, *PASP*, 115, 688
- Kelson, D. D., et al. 2005, in preparation
- Marcy, G. W., & Butler, R. P. 1992, *PASP*, 104, 270
- Marsh, T. R. 1989, *PASP*, 101, 1032
- Mukai, K. 1990, *PASP*, 102, 183
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Cambridge: University Press, 1992, 2nd ed.
- Robertson, J. G. 1986, *PASP*, 98, 1220
- Smith, S. 1936, *ApJ*, 83, 23

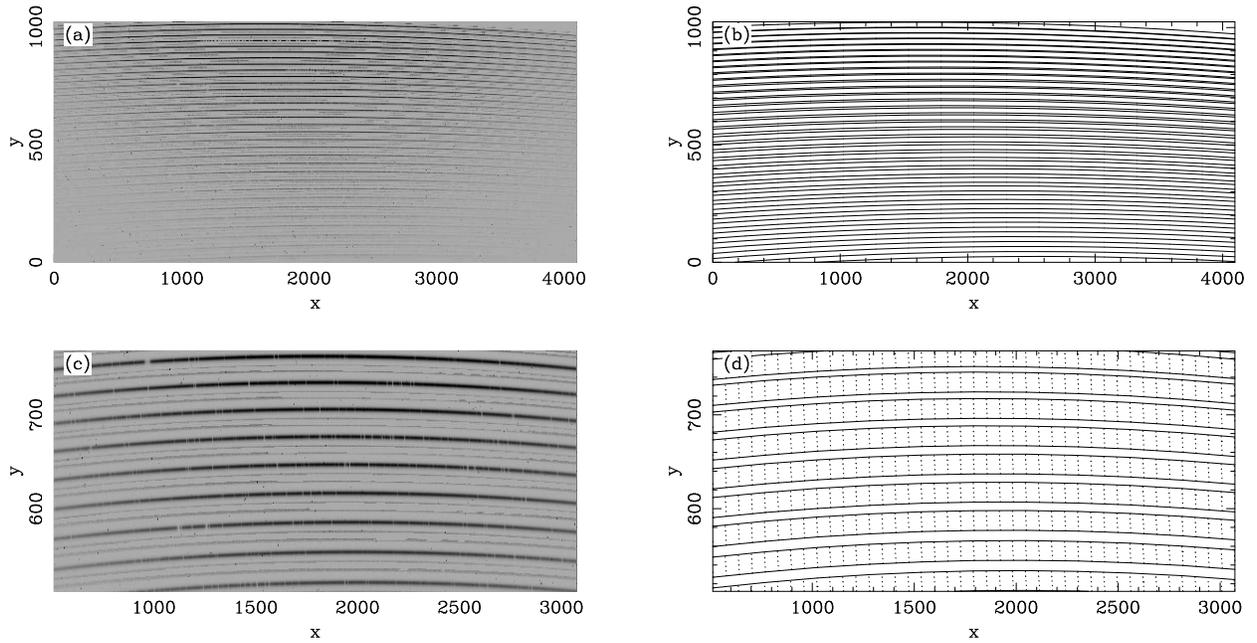


Fig. 1.— A single background-subtracted MIKE exposure. (a) The full frame; (b) contours of constant  $y_t$ . The contours shown are those that correspond to the boundaries of the orders. (c) A subsection of the data; (d) the corresponding section’s contours of constant  $y_t$  (solid lines) and  $x_r$  (dotted lines). Figure 2 shows a subsection of these data, with vectors illustrating the increasing directions of  $y_t$  and  $x_r$ .

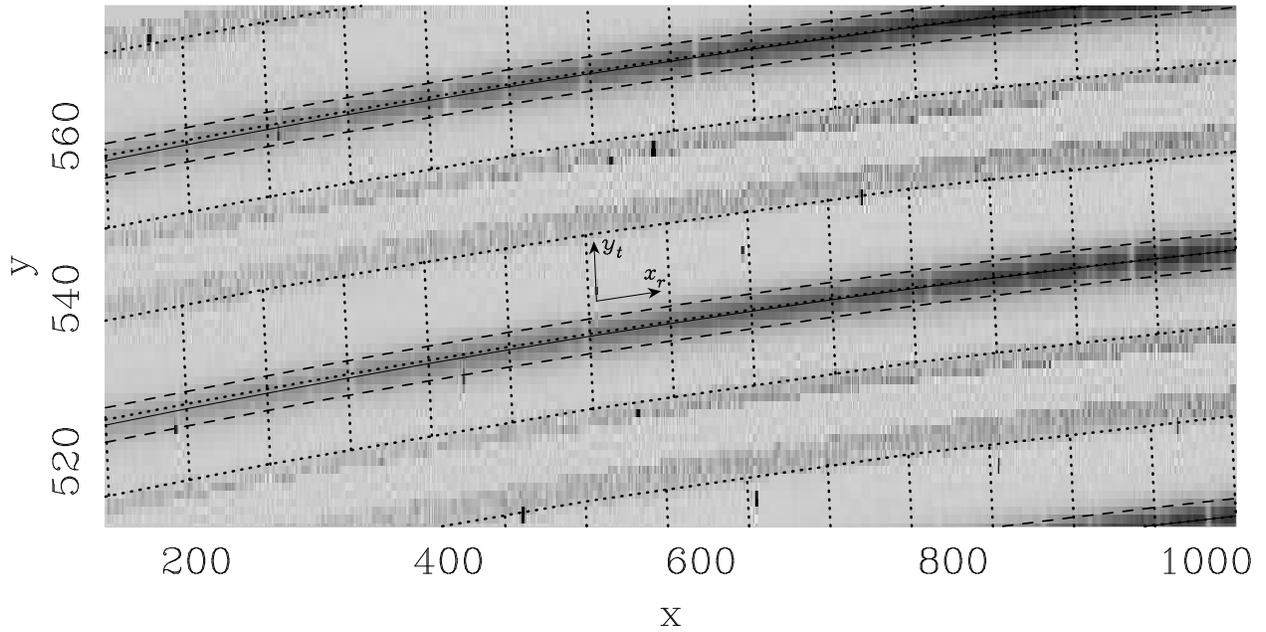


Fig. 2.— A subsection of the data in Figure 1 near  $5800\text{\AA}$ . Contours of constant  $y_t$  and  $x_r$  are shown by the dotted lines. Vectors are also shown to indicate the directions of increasing  $y_t$  and  $x_r$ . The first moment of the object,  $\langle y_t \rangle$ , is shown by the solid line. The first moment is characterized as a function of both  $a$ , the spectral order, and  $x_r$ , the wavelength-dependent coordinate. The locations of the second moment,  $\langle y_c^2 \rangle^{1/2}$ , are shown by the dashed lines on either side of the first moment.

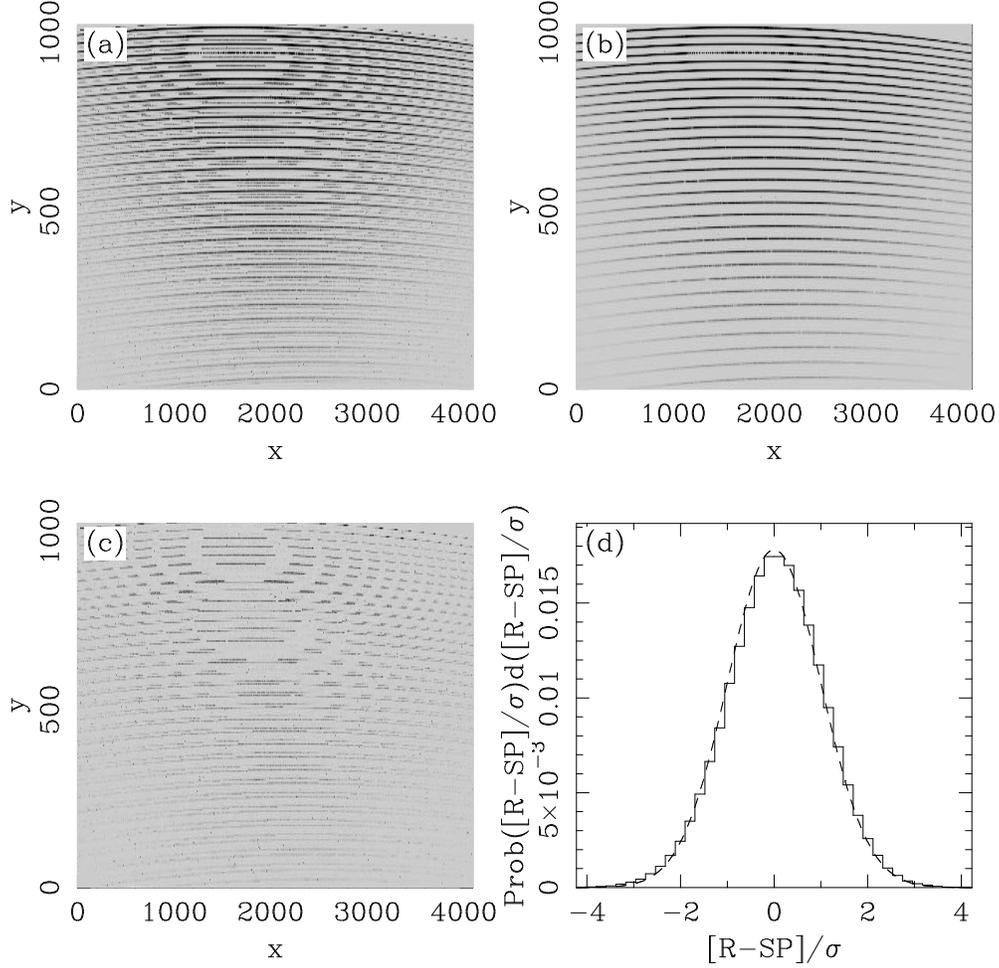


Fig. 3.— A MIKE exposure containing several orders of a high signal-to-noise ratio spectrum of a star. In (a) we show  $R(x, y)$ , the background-subtracted spectrum. In (b) we show  $S(x_r, a_l)P(y_s, a_l)$ , sampled onto the same grid of coordinates over which  $R$  is defined. In (c) we show the residuals from the fit. For these data,  $S$  was constructed as a fit to three exposures. (d) The distribution of residuals of the single exposure with respect to the fit to the three exposures, normalized by the expected noise. The distribution should be well-described by a Gaussian with a standard deviation of unity, as shown by the dashed line. The bi-weight estimators for location and scale (Beers et al. 1990) indicate a  $< 1\%$  deviation from the expected values for the first and second moments of the distribution.

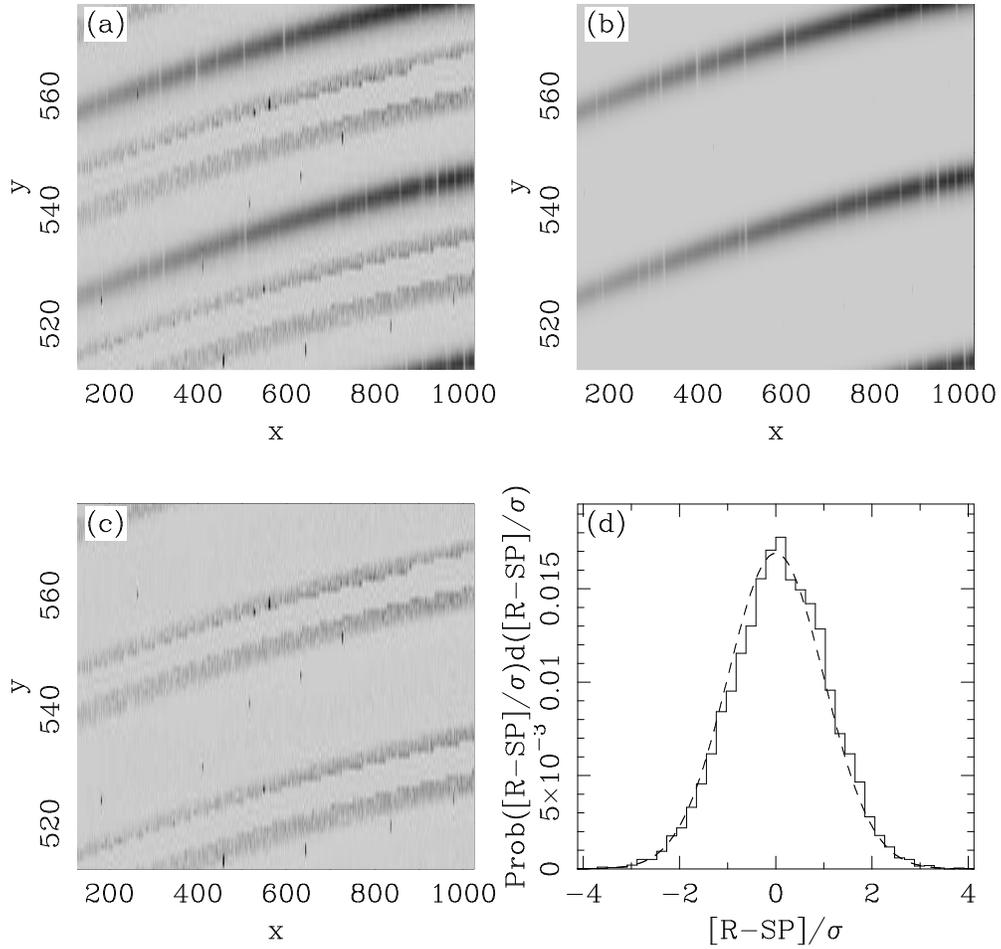


Fig. 4.— Same as in Figure 3 but for a subsection of the exposure (the same section as in Figure 2). The bi-weight estimators for location and scale indicate a 1% deviation from the expected values for the first and second moments of the distribution.

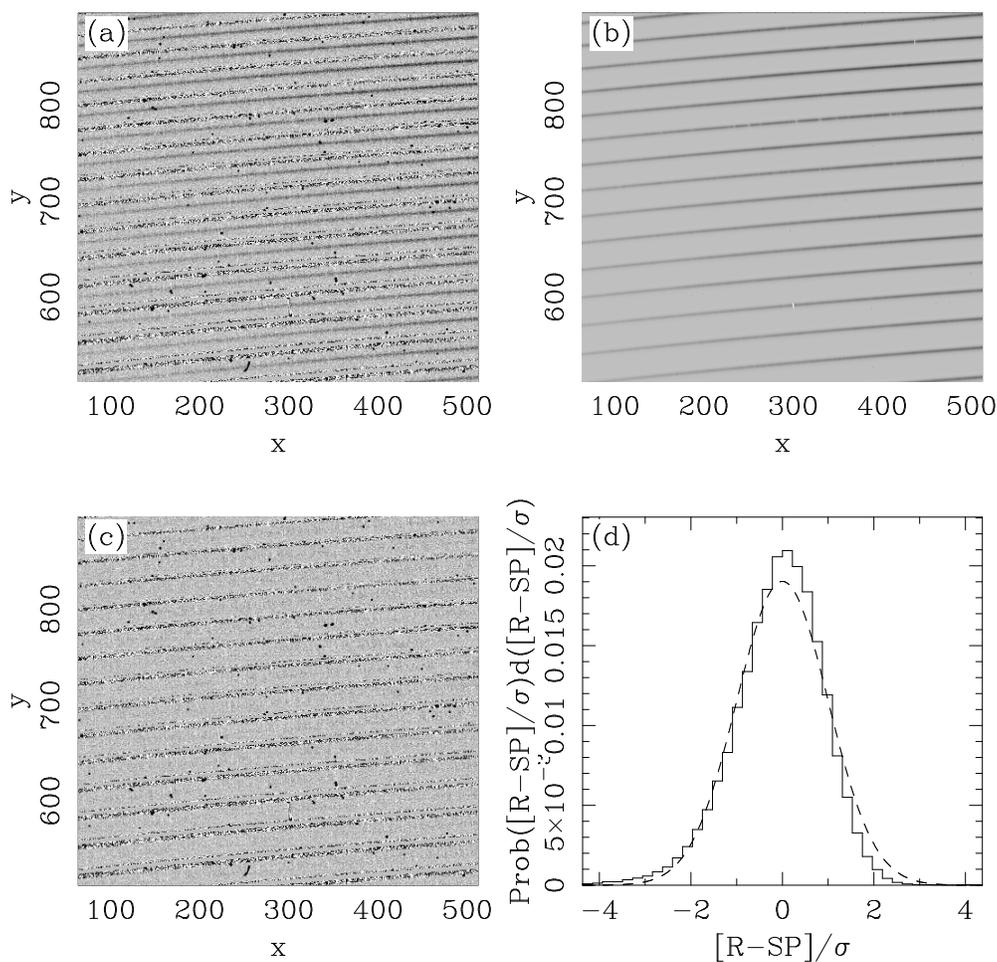


Fig. 5.— Same as in Figure 3 but for a subsection of data on a faint source. For these data the fitting for  $S$  was performed using six exposures. The distribution of residuals is skewed slightly and the small non-Gaussiannity results from noise being dominated by the read noise. The bi-weight estimators for location and scale indicate a 4% deviation from the expected values for the first and second moments of the distribution.

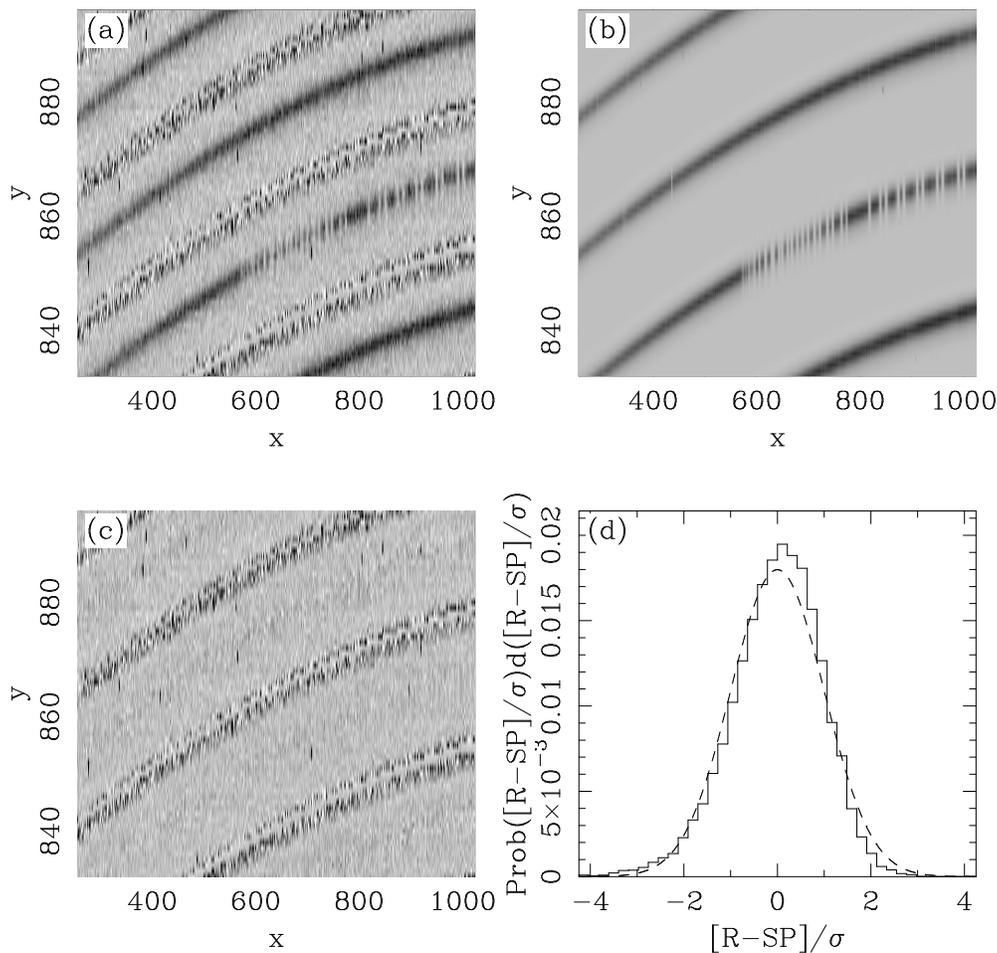


Fig. 6.— Same as in Figure 5 but for a redder subsection of the data, around the region containing the A band. Again, the fit for  $S$  was performed using six exposures. The noise is becoming dominated by that in the object (plus the background) and the distribution of residuals is less skewed than in Figure 5. The bi-weight estimators for location and scale indicate a 3% deviation from the expected values for the first and second moments of the distribution.

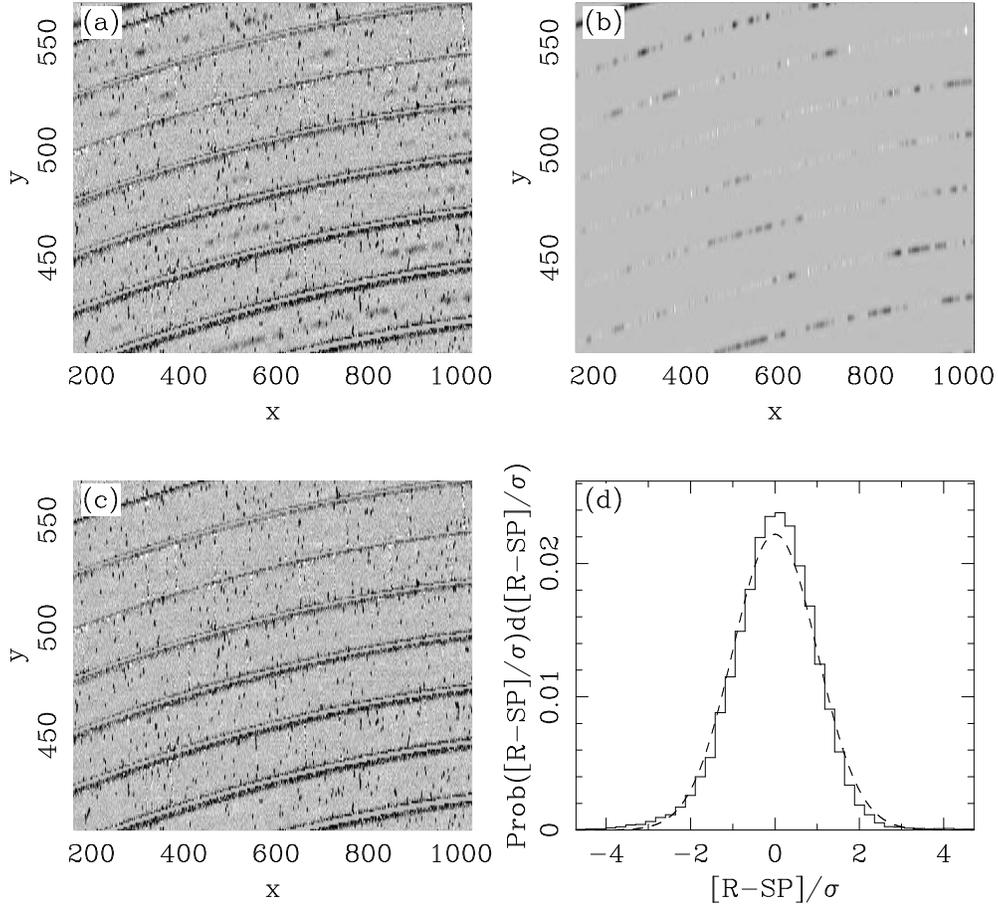


Fig. 7.— (a) A portion of one of four 3000s exposures of a  $z = 5.78$  QSO. The region shows portions of orders from  $7200\text{\AA}$  to  $8200\text{\AA}$ . (b) The fit of  $S(x_r, a_l)P(y_s, a_l)$  over the same region. (c) The residuals from the fit. (d) The distribution of residuals from the fit, normalized by the expected noise. The noise in these data is marginally dominated by the read noise. The bi-weight estimators for location and scale indicate a 1% deviation from the expected values for the first and second moments of the distribution.