



# Modeling the Echelle Spectra Continuum with Alpha Shapes and Local Regression Fitting

Xin Xu<sup>1</sup>, Jessi Cisewski-Kehe<sup>1</sup>, Allen B. Davis<sup>2</sup>, Debra A. Fischer<sup>2</sup>, and John M. Brewer<sup>2</sup>

<sup>1</sup>Department of Statistics and Data Science, Yale University, 24 Hillhouse Ave, New Haven, CT 06511, USA; [xin.xu@yale.edu](mailto:xin.xu@yale.edu)

<sup>2</sup>Department of Astronomy, Yale University, 52 Hillhouse Ave, New Haven, CT 06511, USA

Received 2019 March 3; revised 2019 April 11; accepted 2019 April 20; published 2019 May 30

## Abstract

Continuum normalization of echelle spectra is an important data analysis step that is difficult to automate. Polynomial fitting requires a reasonably high-order model to follow the steep slope of the blaze function. However, in the presence of deep spectral lines, a high-order polynomial fit can result in ripples in the normalized continuum that increase errors in spectral analysis. Here, we present two algorithms for flattening the spectrum continuum. The Alpha-shape Fitting to Spectrum algorithm is completely data driven, using an alpha shape to obtain an initial estimate of the blaze function. The Alpha-shape and Lab Source Fitting to Spectrum algorithm incorporates a continuum constraint from a laboratory source reference spectrum for the blaze function estimation. These algorithms are tested on a simulated spectrum, where we demonstrate improved normalization compared to polynomial regression for continuum fitting. We show an additional application, using the algorithms for mitigation of spatially correlated quantum efficiency variations and fringing in the charge-coupled device detector of the EXtreme PREcision Spectrometer.

**Key words:** instrumentation: spectrographs – methods: data analysis – methods: statistical – techniques: radial velocities – techniques: spectroscopic

## 1. Introduction

Spectroscopy is a powerful observational technique for understanding fundamental astrophysics. A high-dispersion stellar spectrum contains detailed information about individual atomic transitions that enables the derivation of parameters such as effective temperature, surface gravity, and elemental abundances, and the measurement of Doppler shifts in stellar spectra reveals the presence of stellar and planetary companions (Fischer et al. 2016, and references within). Most spectroscopic analysis techniques require a flat, continuum-normalized spectrum (Blanco-Cuaresma et al. 2014). For example, the precision of equivalent width measurements for abundance analysis or cross-correlation for exoplanet detection is very sensitive to even small errors in continuum normalization (Torres et al. 2012, and references within).

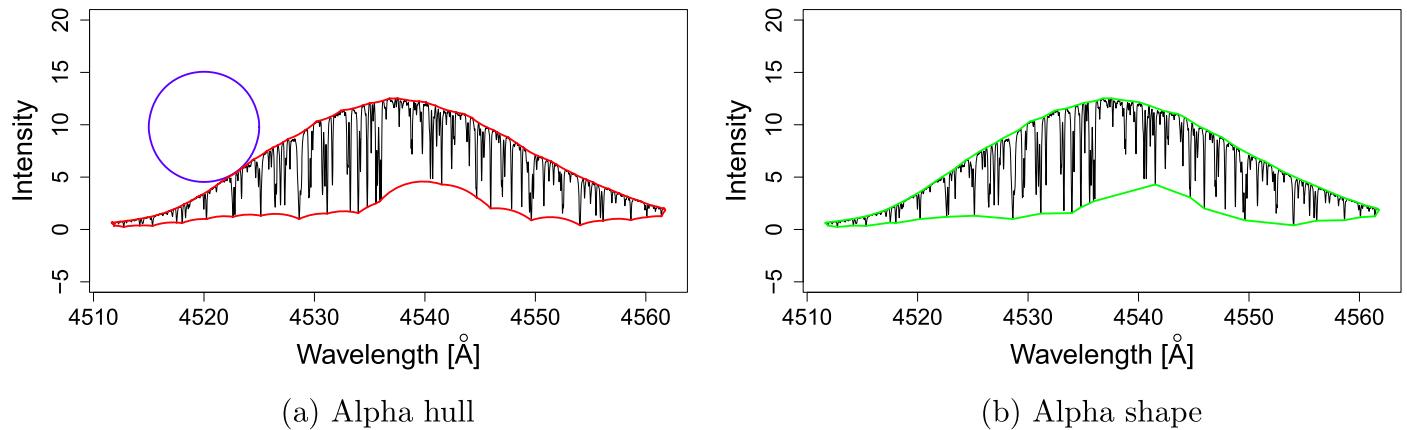
An echelle spectrograph disperses light so that high spectral orders can be recorded. The higher orders have greater dispersion and, therefore, provide higher resolution spectra over a broad range of wavelengths. However, most of the brightness of a spectrum is concentrated in the zeroth and lower spectral orders; higher orders are intrinsically fainter. The optical grating in an echelle spectrograph has angled facets designed to shift the intensity envelope of dispersed light to high spectral orders. When this phase shift is introduced, the grating is said to be blazed, and with cross-dispersion, several dozen spectral orders can be stacked onto the detector. Each order has an intensity distribution characterized by the blaze function so that the continuum intensity is strongest in the center of the order and drops off steeply toward the edges of the order. The term “blaze function” is often used to describe the shape of the continuum across an echelle order, and, indeed, the blaze distribution is the dominant effect.

One approach for flattening the spectrum is to divide by the theoretical blaze function (Barker 1984), which depends on grating parameters (incident angle, the angle of the blazed facets, and the grating facet width and spacing) and can be calculated for each order of the spectrum. However, this method for normalization or flattening of the spectrum will leave residual variations in the continuum because of departures from the theoretical blaze function: manufacturing defects in the grating, chromatic aberrations in the optics of a spectrograph, or quantum efficiency (QE) variations in the electronic detector. In addition to the instrumental blaze, additional wavelength-dependent intensity variations will occur because of the blackbody temperature of stars or calibration lamps.

Another approach for normalizing the continuum is to divide by an extracted flat-field calibration source (e.g., Skoda et al. 2008), such as a quartz lamp. However, the quartz lamp will also have a blackbody curve superimposed on the blaze function. One of the most common methods for normalizing the continuum of a spectrum is to fit a polynomial to high points along the order. This is a completely agnostic approach that does not require prior knowledge about the blaze function, and although it works fairly well, polynomial fitting can fail near broad and deep lines, especially if they are close to the edges of the orders. Here, we describe a new approach for continuum fitting and compare our method with polynomial fitting in Section 3.

## 2. Description of Methods

In this work, the goal is to estimate the blaze function of a target spectrum so that it may be removed, leaving a flattened spectrum. Two algorithms are proposed to accommodate different scenarios: (i) the *Alpha-shape Fitting to Spectrum algorithm* (AFS) (the baseline approach) and (ii) the *Alpha-shape*



**Figure 1.** (a) Alpha hull with  $\alpha = 5$  in red. The blue circle is an example of  $\alpha$ -ball illustrating the process of generating the alpha hull. (b) The resulting alpha shape in green, by straightening the red arcs from (a).

and Lab Source Fitting to Spectrum algorithm (ALSFS) (when a laboratory continuum source is available).<sup>3</sup>

In the proposed algorithms, we first fit an *alpha shape* (Edelsbrunner et al. 1983), which is a polygon enclosing a data set, to capture the general shape of the blaze function of a target spectrum. Then we use this preliminary estimation to select a set of pixels that are ultimately used to fit the final blaze function model. The pixels are selected so that they are generally on or near the continuum and not in the absorption lines. With the selected set of pixels, we use local polynomial regression (Cleveland 1979) to estimate the blaze function. The two algorithms are introduced next, followed by a discussion on how to select the tuning parameters.

### 2.1. AFS Algorithm

The proposed AFS algorithm is a versatile algorithm that can be used to remove the blaze function whether or not a corresponding laboratory source spectrum is available. In this algorithm, an alpha shape is used to obtain a preliminary estimation of the blaze function's high-level shape. An alpha shape is a generalization of a convex hull, but is not required to be a convex set; it is a region bounded by a set of segments generated from a set of points. To generate an alpha shape, imagine a piece of paper with a plot of spectrum printed on it. Consider a special paper cutter that can only cut out circles with radius  $\alpha$ , known as  $\alpha$ -balls. An incomplete circle is allowed, but the cutter cannot cut anything from the spectrum. In Figure 1(a), the blue circle is an  $\alpha$ -ball that can be cut out from the paper, but we cannot move the  $\alpha$ -ball any lower vertically because it would cut the spectrum. Continue to cut as many  $\alpha$ -balls as possible, and the remaining paper is called an *alpha hull*. Figure 1(a) shows the resulting alpha hull with  $\alpha = 5$ . By connecting the points where the alpha hull touches the spectrum with straight segments, it becomes an alpha shape, as displayed in Figure 1(b). An alpha shape can capture the general shape of a spectrum, and its upper boundary is used as a starting model for the blaze function estimation.

Another technique used in the AFS algorithm is local polynomial regression (Cleveland 1979), which is a nonparametric method for estimating functions given a set of points. At each point,  $(\lambda_i, y_i)$ , a  $p$ -degree polynomial model is fitted to a

subset of neighboring points of  $\lambda_i$ . The  $p$ -degree polynomial regression is fitted by weighted least squares, giving more weight to points closer to  $\lambda_i$  and less weight to points further away. Consider a data set  $(\lambda_1, y_1), (\lambda_2, y_2), \dots, (\lambda_n, y_n)$ , and a  $p$ -degree polynomial function in the neighborhood of  $\lambda_i$  is  $f_{\beta^{(\lambda_i)}}(\lambda_j) = \beta_0^{(\lambda_i)} + \beta_1^{(\lambda_i)}(\lambda_j - \lambda_i) + \beta_2^{(\lambda_i)}(\lambda_j - \lambda_i)^2 + \dots + \beta_p^{(\lambda_i)}(\lambda_j - \lambda_i)^p$ , where  $\lambda_j$  is in the neighborhood of  $\lambda_i$ . Then the estimation of  $\beta^{(\lambda_i)}$ ,  $\hat{\beta}^{(\lambda_i)} = (\hat{\beta}_0^{(\lambda_i)}, \hat{\beta}_1^{(\lambda_i)}, \dots, \hat{\beta}_p^{(\lambda_i)})$ , is obtained by minimizing

$$\sum_{j \in N_{m_0}(\lambda_i)} \omega_i(\lambda_j)(y_j - f_{\beta^{(\lambda_i)}}(\lambda_j))^2, \quad (1)$$

where  $N_{m_0}(\lambda_i)$  is the neighboring set of  $\lambda_i$  containing the nearest  $[m_0 n]$  pixels to  $\lambda_i$ ,  $m_0$  is a smoothing parameter, and

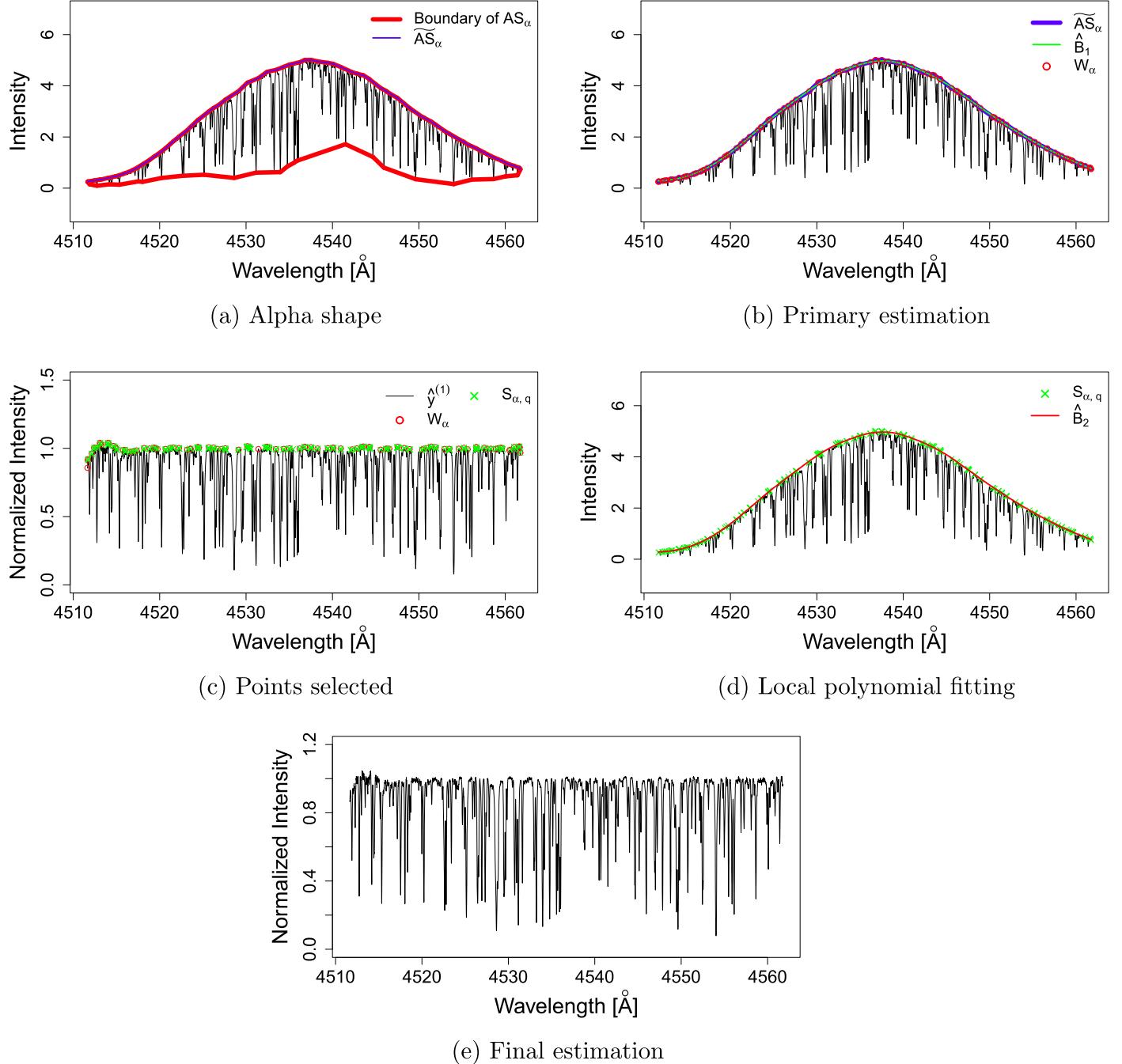
$$\omega_i(\lambda_j) = K \left( \frac{|\lambda_j - \lambda_i|}{\max_{j^* \in N_{m_0}(\lambda_i)} |\lambda_{j^*} - \lambda_i|} \right), \text{ with } K(x) = (1 - x^3)^3 \text{ as a common weighting function.}$$

The local polynomial estimate at  $\lambda_i$  is  $\hat{\beta}_0^{(\lambda_i)}$ . An advantage of local polynomial regression over ordinary polynomial regression is its ability to adapt to local characteristics of a data set rather than fitting all the data points using one single model. This flexibility makes it useful for modeling complex data sets where regular polynomial regression fails.

Let  $\{(\lambda_i, y_i)\}_{i=1}^n$  be an observed spectrum, where  $\lambda_i$  is the wavelength of pixel  $i$ , and  $y_i$  is the intensity of pixel  $i$ . Our method is summarized in Algorithm 1 and illustrated in Figure 2 with the details of the AFS algorithm presented next.

In step 1, the intensity vector,  $y = (y_1, y_2, \dots, y_n)$ , is rescaled by multiplying by a value  $u = \frac{\max(\lambda) - \min(\lambda)}{10 \times \max(y)}$ . This  $u$  corresponds to the  $\alpha$  value for the alpha shape. Because the construction of an alpha shape depends on coverage areas of small circles, the relative range of  $\lambda$  and  $y$  affects results: if the range of  $y$  is too large or too small compared to the range of  $\lambda$ , larger alpha values should be used for the alpha shape. For a common blaze function,  $u = \frac{\max(\lambda) - \min(\lambda)}{10 \times \max(y)}$  works well by scaling the range of  $y$  and  $\lambda$  to be 1: 10, in coordination with a recommended value for  $\alpha$  later. In step 2, the alpha shape,  $AS_\alpha$ , is constructed with radius  $\alpha$ , which is an infinite point set containing all the points within the boundary of the alpha shape (see Figure 2(a)). In  $AS_\alpha$ , for each  $\lambda_i$ , there are infinite  $y_i^*$  such that  $(\lambda_i, y_i^*) \in AS_\alpha$ . The upper boundary of  $AS_\alpha$  is defined as

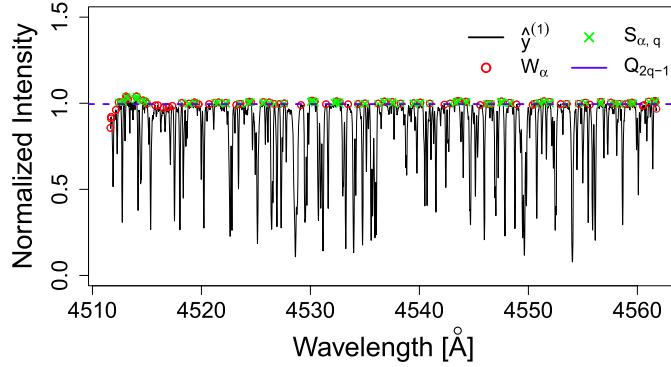
<sup>3</sup> The implementation code of AFS and ALSFS, as well as example data, can be found and downloaded from: <https://github.com/xinxuyale/AFS> or <https://zenodo.org/badge/latestdoi/173169370>.



**Figure 2.** Illustration of the AFS Algorithm 1. (a) Steps 1 and 2: the alpha shape of the whole spectrum  $AS_\alpha$  and its upper boundary  $\widetilde{AS}_\alpha$ . The red arcs represent the boundary of the alpha shape  $AS_\alpha$ , and the blue line connects the points in  $\widetilde{AS}_\alpha$ . (b) Step 3: a smoothed version of  $\widetilde{AS}_\alpha$ , denoted as  $\hat{B}_1$  (in green). The red circles are  $W_\alpha$ : the intersection of  $\widetilde{AS}_\alpha$  and the spectrum  $\{(\lambda_i, y_i)\}_{i=1}^n$ . (c) The last part of step 3, and step 4: divide  $y$  by  $\hat{B}_1$  to get a primary blaze-removed spectrum  $\hat{y}^{(1)}$  (in black) and select points to the set  $S_{\alpha,q}$  (green exes) using  $\hat{y}^{(1)}$ . (d) Step 5: a local polynomial fitting  $\hat{B}_2$  (in red) using points in  $S_{\alpha,q}$  (green exes), which is the final estimation of the blaze function. (e) Step 6: final blaze-removed spectrum.

$\widetilde{AS}_\alpha$ , which is a finite point set only including the largest  $y_i^*$  such that  $(\lambda_i, y_i^*) \in AS_\alpha$  for each  $\lambda_i$ , as displayed in Figure 2(a). In step 3, we fit a local polynomial regression using all the points in  $\widetilde{AS}_\alpha$ , denoted as  $\hat{B}_1$ , such that  $\hat{B}_1$  is a smoothed version of  $\widetilde{AS}_\alpha$  and is the initial model of the blaze function (see Figure 2(b)).  $\hat{B}_1$  is not an accurate estimate of the blaze function, but rather an approximation of its shape. Next,  $y$  is divided by  $\hat{B}_1$  to get the first estimate of the flattened spectrum, denoted as  $\hat{y}^{(1)}$ , displayed in Figure 2(c).

In step 4, we denote the intersection of  $\widetilde{AS}_\alpha$  and the spectrum  $\{(\lambda_i, y_i)\}_{i=1}^n$  as  $W_\alpha$ , shown in Figures 2(b) and (c). Notice that  $W_\alpha$  contains points mostly near the continuum. The blue line in Figure 2(b), which connects the points in  $\widetilde{AS}_\alpha$ , can be thought of as a collection of segments, and points in  $W_\alpha$  are vertices of those segments. Each point in  $W_\alpha$  is a vertex of a window where splits are defined—the spectrum is cut into small windows by the points in  $W_\alpha$  in order to get the local quantiles. The purpose of the next steps is to select a subset of points that



(a) Points selected

**Figure 3.** (a) Divide  $y$  by  $\hat{B}_1$  to get  $\hat{y}^{(1)}$  (in black). Select points to the set  $S_{\alpha,q}$  (green exes) both locally and globally using  $\hat{y}^{(1)}$ . The blue dashed line is the global quantile  $Q_{2q-1}$ . (b) The blue dashed line is the original laboratory source spectrum. Over the limited wavelength range of an order, the effective blackbody function is approximately linear, and so the laboratory source spectrum generally has the same shape as the true blaze function but needs linear modifications: intercept, scale, and slope parameters are used in the linear transformation to get  $\hat{B}_2$ . The red solid line is our final estimate for the blaze function.

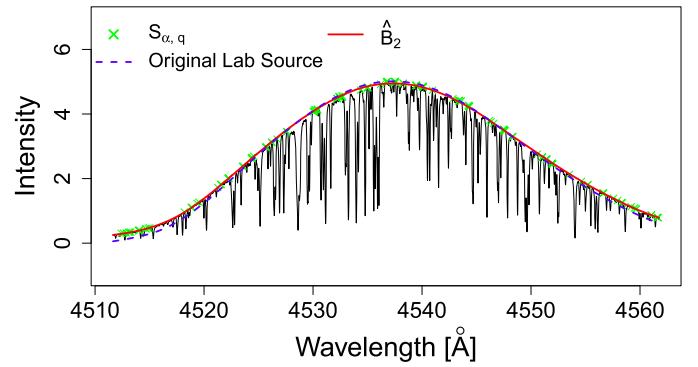
do not fall into an absorption feature. This is accomplished by selecting  $\hat{y}_i^{(1)}$  in the upper quantiles of these windows. In particular, in the  $j$ th window, we select the points whose  $\hat{y}_i^{(1)}$  values are larger than the  $q$  quantile of the  $\hat{y}_i^{(1)}$  in the window, and the selected points make up the set  $S_{j,\alpha,q}$ . For  $j = 1, 2, \dots, |W_\alpha| - 1$ , where  $|W_\alpha|$  is the number of points in  $W_\alpha$ , the combined set of  $S_{j,\alpha,q}$  for different  $j$ 's is defined as  $S_{\alpha,q}$ , displayed in Figure 2(c).  $S_{\alpha,q}$  contains points that are locally in the upper  $1 - q$  quantile, which will be used to estimate the blaze function because these points generally do not fall into an absorption line. In step 5, we run a local polynomial regression on  $S_{\alpha,q}$  and fit it to the whole spectrum. This regression is our final estimate of the blaze function, denoted as  $\hat{B}_2$ . The red line in Figure 2(d) shows the final estimate of the blaze function. Then, in step 6,  $y$  is divided by  $\hat{B}_2$  to get the blaze-removed spectrum, denoted as  $\hat{y}^{(2)}$  and shown in Figure 2(e).

#### Algorithm 1. AFS Algorithm

- 
- Step 0: Let  $\{(\lambda_i, y_i)\}_{i=1}^n$  be an observed spectrum.  
 Step 1: Let  $u = \frac{\max(\lambda) - \min(\lambda)}{10 \times \max(y)}$ . Multiply  $y$  by  $u$ .  
 Step 2: Let  $\text{AS}_\alpha = \text{alpha\_shape}(\{(\lambda_i, y_i), i = 1, \dots, n\})$  with radius value  $\alpha$ .  
 Then  
 $\widetilde{\text{AS}}_\alpha = \{(\lambda_i, \tilde{y}(\lambda_i)): \lambda_i \in \{\lambda_i, i = 1, \dots, n\}, \tilde{y}(\lambda_i) = \max_{\forall (\lambda_j, y_j^*) \in \text{AS}_\alpha} y_j^*\}$ .  
 Step 3: Run a local polynomial regression on  $\widetilde{\text{AS}}_\alpha$  with smoothing parameter  $m_0$ , denoting the fit model as  $\hat{B}_1$ . Calculate  $\hat{y}^{(1)} = \frac{y}{\hat{B}_1}$ .  
 Step 4: Let  $W_\alpha = \widetilde{\text{AS}}_\alpha \cap \{(\lambda_i, y_i)\}_{i=1}^n = \{(\lambda_i, y_i), i = w_1, w_2, \dots, w_{|W_\alpha|}\}$ . Let  
 $S_{j,\alpha,q} = \left\{ w_j \leq i \leq w_{j+1}: \frac{\sum_{k=w_j}^{w_{j+1}} \mathbb{1}(\hat{y}_k^{(1)} \geq \hat{y}_i^{(1)})}{w_{j+1} - w_j + 1} \geq q \right\}$ .  
 The  $S_{\alpha,q} = \bigcup_{j=1, \dots, |W_\alpha| - 1} S_{j,\alpha,q}$ .  
 Step 5: Run a local polynomial regression on set  $\{(\lambda_i, y_i)\}_{i \in S_{\alpha,q}}$  with  $m_0$  and fit to the whole spectrum, denoted as  $\hat{B}_2$ .  
 Step 6: Calculate  $\hat{y}^{(2)} = \frac{y}{\hat{B}_2}$ . Output  $\{(\lambda_i, \hat{y}_i^{(2)})\}_{i=1}^n$ .
- 

#### 2.2. ALSFS Method

The ALSFS algorithm is a method for removing the blaze function when a laboratory source spectrum, such as an LED or



(b) Local polynomial fitting

quartz lamp spectrum, is available as a reference. Generally, when a reliable reference spectrum is available, it can be used as the preliminary estimate of the blaze function shape. The use of a reference spectrum can be particularly advantageous in situations when the science spectrum contains wide absorption lines, which often make estimation of blaze function shape especially challenging.

The AFS algorithm in Section 2.1 can be adapted to take advantage of this additional information in the reference spectrum. To a first approximation, a laboratory source continuum spectrum should trace the instrumental blaze function for each order. However, the calibration source will typically have some effective blackbody temperature—that is, its intrinsic intensity will peak at a specific wavelength. Over the limited wavelength range covered by a single spectral order, this effective blackbody function is approximately linear. Therefore, an observation of a laboratory source spectrum can be modeled as a linear transformation of the instrumental blaze function, and the difference between the blaze function and the corresponding laboratory source spectrum is only a location-scale transformation. Thus, the blaze estimation problem can be translated into an optimization problem to find the best intercept, scale, and slope of the laboratory source.

#### 2.2.1. ALSFS Algorithm

The ALSFS algorithm is initialized in the same manner as the AFS algorithm in steps 1 and 2, but it differs slightly in steps 3 and 4 and greatly in step 5. Let the reference laboratory source spectrum be  $\{(\lambda_i, l_i)\}_{i=1}^n$ . The ALSFS algorithm is summarized in Algorithm 2.

Steps 1 and 2 are the same as the AFS algorithm. Because prior knowledge of the shape of blaze function is available, fewer points are needed for the second local polynomial fitting. Step 3 is similar to step 3 of the AFS algorithm, but we make set  $S_{\alpha,q}$  more accurate by also calculating the  $2q - 1$  quantile of  $\hat{y}^{(1)}$ , denoted as  $Q_{2q-1}$ .  $Q_{2q-1} = Q_{1-2(1-q)}$ , which means the upper  $2(1 - q)$  quantile of  $\hat{y}^{(1)}$ , displayed in Figure 3(a). Compared to the upper  $1 - q$  quantile of  $\hat{y}^{(1)}$  within each window, a smaller quantile is used for the global quantile; otherwise, too many points will be excluded if the same quantile is used. Step 4 also departs from the AFS algorithm: for the  $j$ th window, we select the points  $\lambda_i$  where both  $\hat{y}_i^{(1)} \geq Q_{2q-1}$  and  $\hat{y}_i^{(1)} \geq Q_{1-2(1-q)}$ .

$q$  quantile of the window and  $\hat{y}_i^{(1)} \geq Q_{2q-1}$  into set  $S_{\alpha,q}$ , shown in Figure 3(a). In step 5, we use the laboratory source spectrum's intensity curve as a reference model, displayed in Figure 3(b), to find the best linear coefficients for blaze function estimation using the points selected in the previous step. We apply a linear transformation on the laboratory source spectrum:  $\hat{l}_i(a, b, c) = a + bl_i + cl_i$ , where  $a$ ,  $b$ , and  $c$  are intercept, scale, and slope parameters, respectively. Because our ultimate goal is to have a flat spectrum without the blaze function, we use an objective function  $\sum_{i \in S_{\alpha,q}} \left( \frac{l_i}{\hat{l}_i(a, b, c)} - 1 \right)^2$ , which measures the total distances from the removed spectrum to the constant 1. We minimize this objective function on the set  $S_{\alpha,q}$  to get estimates for  $a$ ,  $b$ , and  $c$ . Then the modified laboratory source spectrum  $\hat{a} + \hat{b}l_i + \hat{c}\lambda_i$  is our final estimate for blaze function, displayed in Figure 3(b). Step 6 is the same as the AFS algorithm,  $y$  is divided by  $\hat{B}_2$  to get the blaze-removed spectrum.

### Algorithm 2. ALSFS Algorithm

---

Step 0: Let  $\{(\lambda_i, y_i)\}_{i=1}^n$  be an observed spectrum, and  $\{(\lambda_i, l_i)\}_{i=1}^n$  be the corresponding laboratory source.  
 Step 1: Let  $u = \frac{\max(\lambda) - \min(\lambda)}{10 \times \max(y)}$ . Multiply  $y$  by  $u$ .  
 Step 2: Let  $\text{AS}_\alpha = \text{alpha shape}(\{(\lambda_i, y_i), i = 1, \dots, n\})$  with radius value  $\alpha$ .  
 and  $\widetilde{\text{AS}}_\alpha = \{(\lambda_i, \tilde{y}(\lambda_i)): \lambda_i \in \{\lambda_i, i = 1, \dots, n\}, \tilde{y}(\lambda_i) = \max_{\forall (\lambda_j, \tilde{y}_j^*) \in \text{AS}_\alpha} y_j^*\}$ .  
 Step 3: Run a local polynomial regression on  $\widetilde{\text{AS}}_\alpha$  with  $m_0$ , denoted as  $\hat{B}_1$ . Calculate  $\hat{y}^{(1)} = \frac{y}{\hat{B}_1}$ . Denote  $Q_{2q-1} = \text{quantile}(\hat{y}^{(1)}, 2q - 1)$ .  
 Step 4: Let  $W_\alpha = \widetilde{\text{AS}}_\alpha \cap \{(\lambda_i, y_i)\}_{i=1}^n = \{(\lambda_i, y_i), i = w_1, w_2, \dots, w_{|W_\alpha|}\}$ . Let  $S_{j,\alpha,q} = \left\{ w_j \leq i \leq w_{j+1}: \frac{\sum_{k=w_j}^{w_{j+1}} \mathbb{1}(\hat{y}_i^{(1)} \geq \hat{y}_k^{(1)})}{w_{j+1} - w_j} \geq q \text{ and } \hat{y}_i^{(1)} \geq Q_{2q-1} \right\}$ .  
 $S_{\alpha,q} = \bigcup_{j=1, \dots, |W_\alpha|-1} S_{j,\alpha,q}$ .  
 Step 5: Consider a linear transformation:  $\hat{l}_i(a, b, c) = a + bl_i + cl_i$ ,  $i = 1, \dots, n$ .  $(\hat{a}, \hat{b}, \hat{c}) = \text{argmin}_{a,b,c} \sum_{i \in S_{\alpha,q}} \left( \frac{l_i}{\hat{l}_i(a, b, c)} - 1 \right)^2$ .  
 $\hat{B}_2 = \hat{a} + \hat{b}l_i + \hat{c}\lambda_i$ ,  $i = 1, \dots, n$ .  
 Step 6: Calculate  $\hat{y}^{(2)} = \frac{y}{\hat{B}_2}$ . Output  $\{(\lambda_i, \hat{y}_i^{(2)})\}_{i=1}^n$ .

---

#### 2.2.2. Laboratory Source Smoothing by AFS Algorithm

The process of flat-fielding spectra is necessary for removing pixel-to-pixel QE variations in charge-coupled devices that are used as detectors in astronomical spectrographs. In the case of fiber-fed echelle spectrographs, flat-fielding can be carried out by extracting a featureless calibration spectrum and dividing the extracted science spectrum order-by-order. However, the flat-field source is generally not perfectly uniform in intensity over a large wavelength range and, therefore, will typically have an effective blackbody temperature that does not match the stellar effective temperature. As a result, this division leaves behind residual trends. By using the AFS algorithm, a model can be fitted to each order of the flat-field calibration spectrum. Division of the flat-field echelle orders by this fitted model will then produce a normalized spectrum that can be used to divide out the QE variation. Figure 4 shows an example of how this process was used to create a normalized flat in red orders that exhibit fringing from interference of red wavelengths in thinned silicon detectors. This fringing can be removed by dividing stellar spectra with this normalized flat.

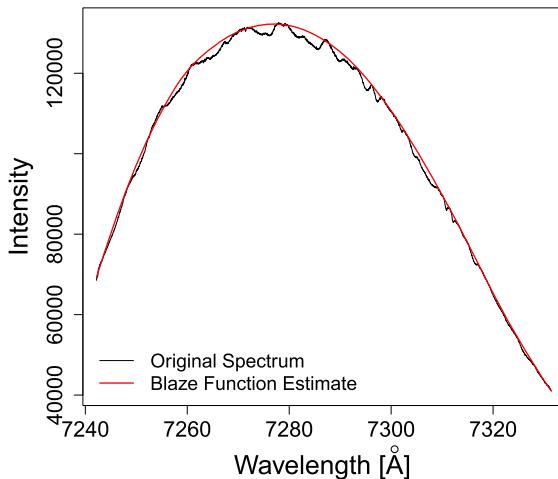
In some orders, a cosmic ray or a pixel with very low QE will create an upward or downward spike that is confined to one or a few pixels. In this case, the AFS algorithm is slightly modified to iteratively reject these pixels using outlier rejection. Let the original laboratory source spectrum be  $\{(\lambda_i, L_i)\}_{i=1}^n$ , and  $\Delta L = \{|L_i - L_{i-1}|, i = 2, \dots, n\}$ . Let  $Q_{q_s}$  be the  $q_s$  quantile of  $\Delta L$ , where  $q_s$  could be a number between 0.95 and 0.99. In the beginning, the 0.99 quantile of  $\Delta L$  is denoted as  $Q_{0.99}^{(0)}$ . In the  $j$ th iteration, remove pixels where  $|L_i - L_{i-1}| > Q_{0.99}^{(j-1)}$  and calculate the 0.99 quantile of the new  $\Delta L$ , denoted as  $Q_{0.99}^{(j)}$ . Continue the iteration until  $Q_{0.99}^{(j)} < Q_{q_s}$ . The remaining pixels are used for smoothing by the AFS algorithm, displayed in Figure 5.

#### 2.3. Parameter Selection

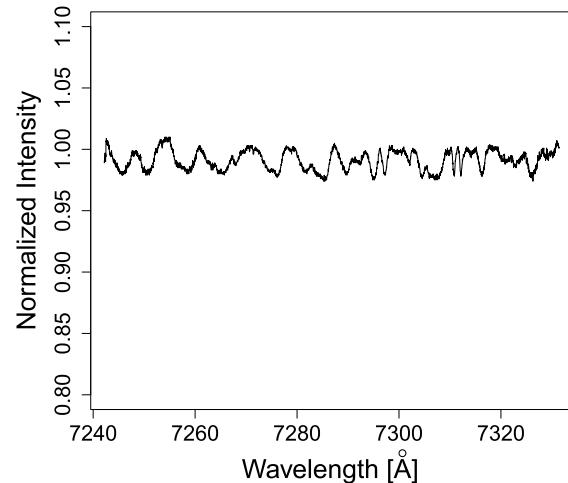
In the proposed algorithms, there are several parameters that need to be selected by users. In the AFS algorithm, there are three parameters:  $\alpha$  for the alpha shape, quantile  $q$  for the point selection, and  $m_0$  in two local polynomial regressions. The  $\alpha$  determines how many windows a spectrum is cut into, because the number of windows is determined by smoothness of the alpha shape, which is controlled by  $\alpha$ . Using these windows,  $q$  determines how many points are selected in each window. After selecting the points,  $m_0$  determines the smoothness of local polynomial fitting on these selected points. Practically, the three parameters are robust within appropriate ranges. For the example spectrum in Figure 6(a), we recommend a set of standard parameters ( $\alpha = \frac{1}{6} \times$  wavelength range,  $q = 0.95$ ,  $m_0 = 0.25$ ) as a default. In Figures 6–8, we show the blaze estimates using extreme parameter choices (top panels) and its resulting normalized spectra compared with the standard parameter choice (bottom panels).

First, the  $\alpha$  can be selected according to the shape of a blaze function. For example, in Figure 2(a), we find that the blaze function increases first and then decreases. Also, it is convex first, then becomes concave, and turns to be convex in the end. We want to select an  $\alpha$  that can capture the convex parts (not too large), but will not go too deep into absorption lines (not too small). The choice of  $\alpha$  is mainly determined by the shape, e.g., curvature and concavity, of a blaze function. However, when there is a wide absorption line, a larger  $\alpha$  may be needed than the shape would suggest. For echelle spectra orders, because each convex portion generally takes up about one-sixth of an order, we recommend selecting an  $\alpha$  that is one-sixth of the wavelength range of the order, but have found empirically that  $\alpha$  is rather robust from  $\frac{1}{12}$  of the wavelength range to one-third of the wavelength range. Figure 6(a) shows an example of a very large  $\alpha$  equal to the order's entire wavelength range. This  $\alpha$  value is too large for the  $\alpha$ -balls to capture the shape of the spectrum near the boundaries; in the blaze-removed spectrum, the left part drops downward like a wide absorption feature because points in that portion of the spectrum were not selected in  $S_{\alpha,q}$ . Figure 6(b) shows an example of a small  $\alpha$  of  $\frac{1}{50}$  of the wavelength range. With such a small  $\alpha$ , the  $\alpha$ -balls fit into absorption lines so that points inside an absorption are selected into set  $S_{\alpha,q}$ . In the removed spectrum, there are regions that are above the reference line  $y = 1$ , compared with the cyan spectrum where  $\alpha = \frac{1}{6} \times$  wavelength range.

The parameter  $q$  depends on the signal-to-noise ratio (S/N) of a spectrum and the amount of absorption. The goal when

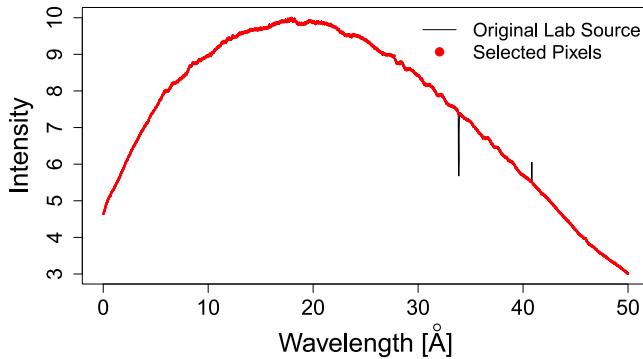


(a) Fringing spectrum



(b) Normalized spectrum

**Figure 4.** (a) Echelle order of the spectrum shows fringing because the thinned silicon charge-coupled device has a thickness comparable to red wavelengths. The AFS algorithm is used to fit a smooth function across the order and division of this order of flux from the flat-field lamp produces a normalized spectrum (b). For very stable spectrographs, the stellar spectra can be divided by this normalized flat-field flux to remove fringing.



**Figure 5.** In the smoothing process, we first use an iteration to get rid of the spikes. In the  $j$ th iteration, remove pixels where  $|L_i - L_{i-1}| > Q_{0.99}^{(j-1)}$  and calculate the 0.99 quantile of the new  $\Delta L$ , denoted as  $Q_{0.99}^{(j)}$ . Iterate until  $Q_{0.99}^{(j)} < Q_s$ . The red points are pixels left to be used for the AFS algorithm.

selecting  $q$  is to find points on the spectrum that do not drop into absorption lines, but instead are on the continuum. After flattening the spectrum using a preliminary estimate of its shape, we select points in the local upper  $1-q$  quantile for the set  $S_{\alpha,q}$  to be used in the final estimation. If we happened to know the true blaze, then after dividing the spectrum by the blaze we would expect to see points randomly scattered around 1. In the proposed algorithms, these points are approximated by set  $S_{\alpha,q}$ . If the S/N is high or there is a large amount of absorption, a larger  $q$  is needed to select points in  $S_{\alpha,q}$  so that these points do not fall in absorption lines. Conversely, if S/N is low or there is minimal absorption, a smaller  $q$  is needed to get enough points into set  $S_{\alpha,q}$ .<sup>4</sup> For an order with a similar amount of absorption as the one displayed in Figure 6(a), a  $q$  from 0.95 to 0.99 works for S/N 300, a  $q$  from 0.85 to 0.95 works for S/N 150, and a  $q$  from 0.5

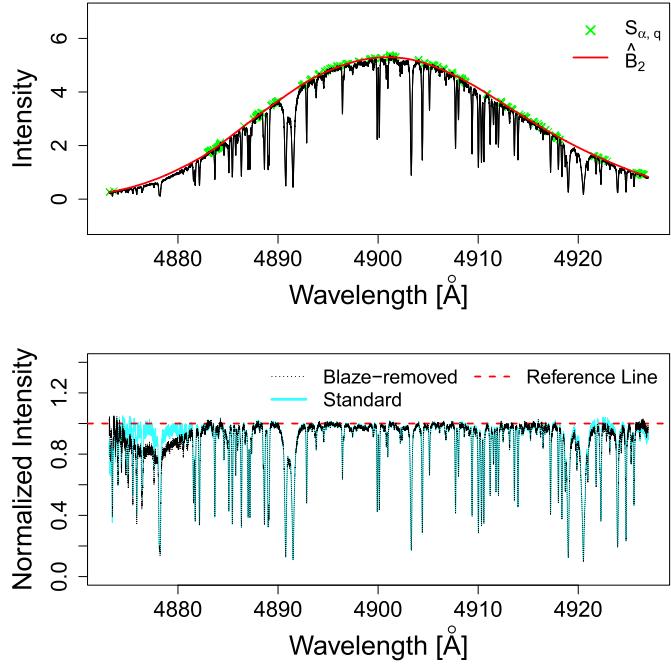
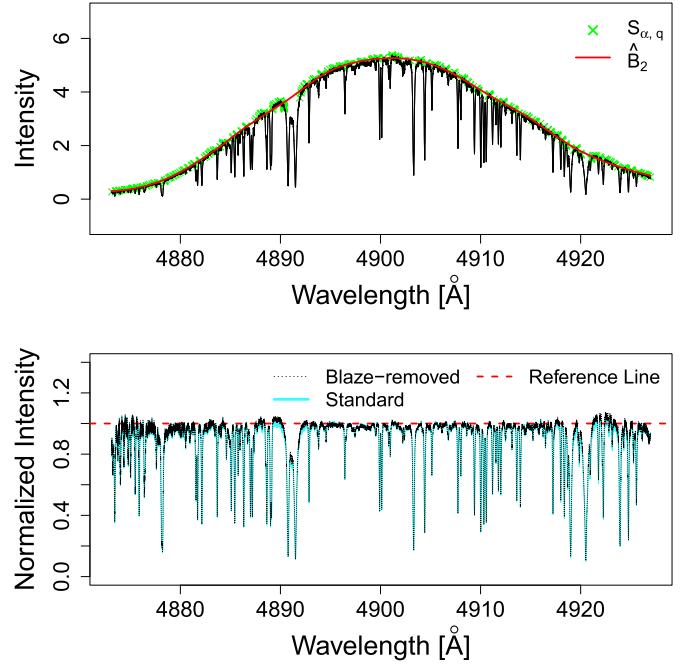
to 0.85 works for S/N lower than 150. Figure 7(a) shows the effects of a small  $q$  such as 0.7: too many points are selected into  $S_{\alpha,q}$  so that the blaze function estimate is dragged downward by the points in absorption lines, and in the removed spectrum, the left and right boundary regions are above the reference line  $y = 1$ . In contrast, Figure 7(b) shows a large  $q$  of 0.999. In this example  $S_{\alpha,q}$  contains too few points, which makes the estimate sit almost completely above the spectrum. In the removed spectrum, almost all the pixels are under the reference line.

The smoothing parameter  $m_0$  depends on the distribution of points in  $S_{\alpha,q}$  along the spectrum, which is determined by the amount of absorption. If there are many absorption lines or any absorption lines that are wide, the set  $S_{\alpha,q}$  has large gaps between pixels, and, thus, a large  $m_0$  is needed to get a good estimate. If there are few absorption lines or absorption lines that are narrow, a small  $m_0$  is needed so that the estimation better adapts to local regions. For an order with a similar amount of absorption as the one displayed in Figure 6(a), an  $m_0$  value from 0.15 to 0.3 has worked well empirically. In Figure 8(a),  $m_0$  is set to be 0.5 (too large) and the estimate is off on the left part of the spectrum, where the blaze-removed spectrum rises to above 1.5. Figure 8(b) shows the results of a too small  $m_0$  value of 0.1: the blaze estimate has some small bumps, but the blaze-removed spectrum looks reasonable to the eye. However, because the true blaze function does not have small bumps, this blaze estimate is not as good a fit as it might appear at first glance.

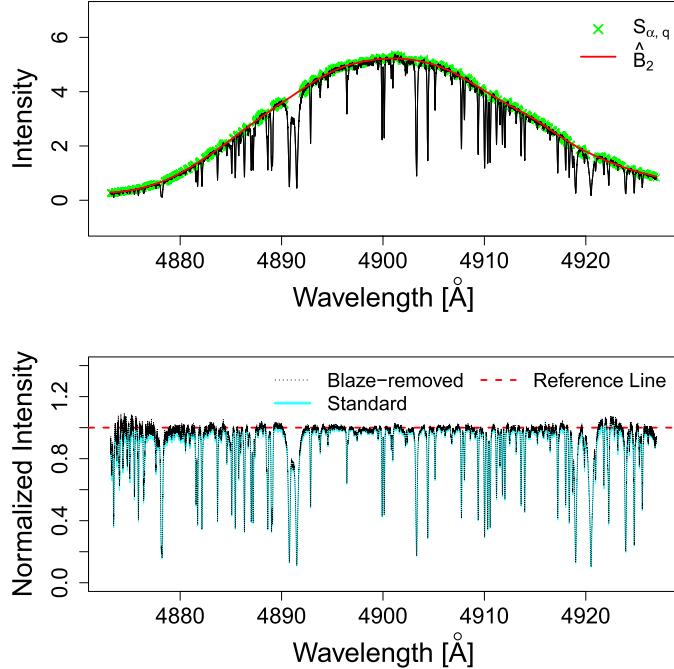
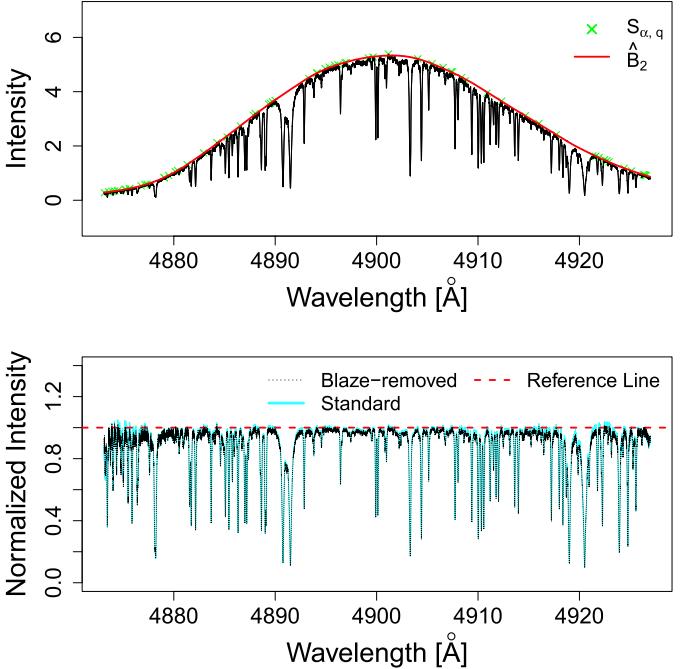
In general, the blaze function estimate is more sensitive to small changes in  $q$  than to  $\alpha$  and  $m_0$ . For an echelle spectrum order with a similar amount of absorption as the one displayed in Figure 6(a), we can start with an  $\alpha$  equal to one-sixth of the wavelength range of the order, an  $m_0$  equal to 0.25, and tune the parameter  $q$  within the range according to its S/N and amount of absorption. A more detailed set of recommendations for parameters in different situations is provided in the Appendix.

The ALSFS algorithm has the same parameters:  $\alpha$ ,  $q$ , and  $m_0$ . Because the final estimate is the reference laboratory source spectrum with linear modification,  $m_0$  has less influence on the results than for the AFS algorithm. In the laboratory source

<sup>4</sup> Alternatively, an adaptive  $q$  can be used for each small window to capture the noise more accurately. An adaptive  $q$  can incorporate the overall S/N and the average intensity in each window to characterize the noise variance better.

(a)  $\alpha=1 \times$  wavelength range(b)  $\alpha=\frac{1}{50} \times$  wavelength range

**Figure 6.** Comparison of the results of the AFS algorithm using extreme values for the  $\alpha$  parameter. (a) Large  $\alpha$ :  $1 \times$  wavelength range. (b) Small  $\alpha$ :  $\frac{1}{50} \times$  wavelength range. The cyan spectrum shows the results from the standard value of  $\alpha = \frac{1}{6} \times$  wavelength range.

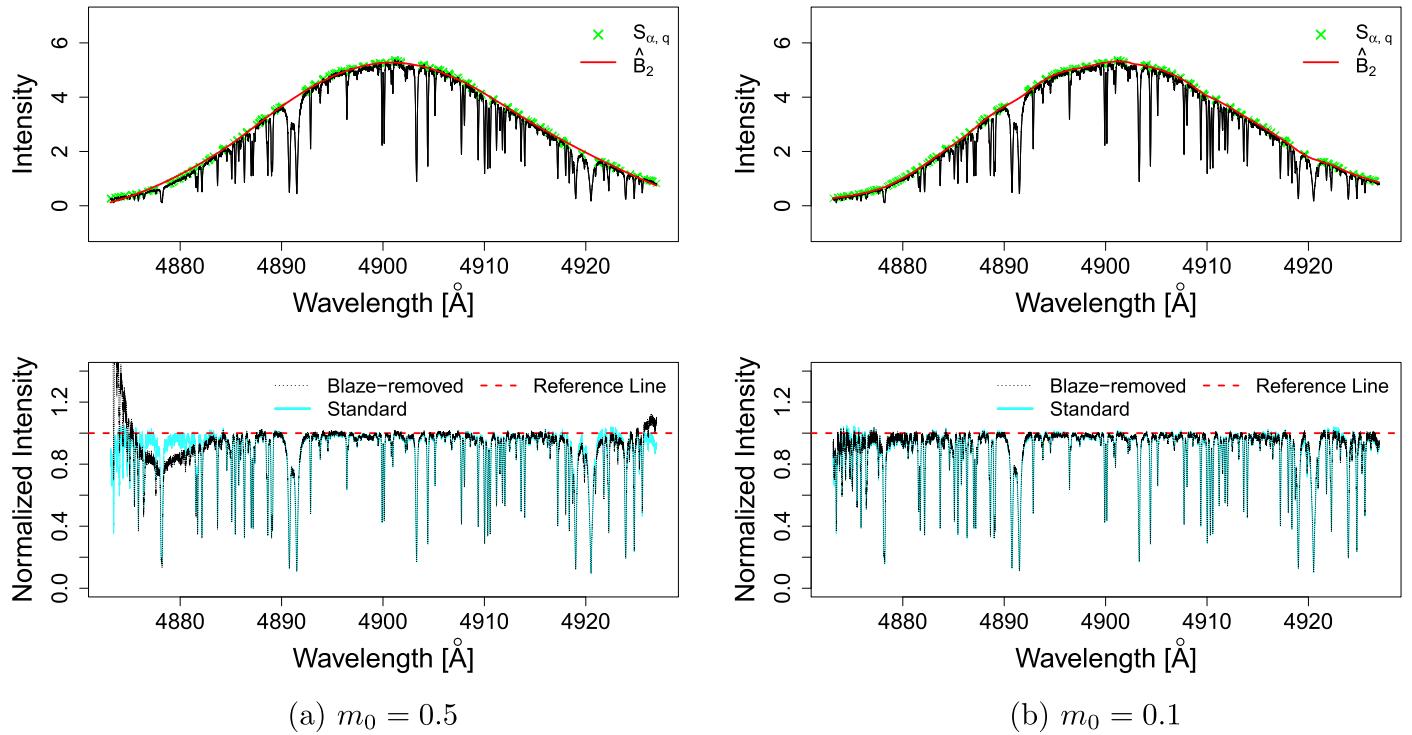
(a)  $q = 0.7$ (b)  $q = 0.999$ 

**Figure 7.** Comparison of the results of the AFS algorithm using extreme values for the  $q$  parameter. (a) Small  $q$ : 0.7. (b) Large  $q$ : 0.999. The cyan spectrum shows the results from the standard value of  $q = 0.95$ .

spectrum smoothing process,  $\alpha$ ,  $q$ , and  $m_0$  operate the same as in the AFS algorithm. The quantile parameter  $q_s$  in  $Q_{q_s}$  depends on the particular appearances of spikes. If spikes are long, use a smaller  $q_s$  such as 0.95; if spikes are very small, use a larger  $q_s$  such as 0.98 or 0.99.

#### 2.4. Complications and Corrections

We have found that the proposed algorithms work well in most cases. However, there are several special cases that can result in poorer estimates of the blaze function, for which we have developed corrections to mitigate these issues. Because



**Figure 8.** Comparison of the results of the AFS algorithm using extreme values for the  $m_0$  parameter. (a) Large  $m_0$ : 0.5. (b) Small  $m_0$ : 0.1. The cyan spectrum shows the results from the standard value of  $m_0 = 0.25$ .

the AFS algorithm relies only on the science spectrum itself, it is more susceptible to complications than the ALSFS algorithm.

#### 2.4.1. Boundary Correction

An order is normalized by dividing by its estimated blaze function. Because the blaze function approaches zero near the edges of the order, small errors in the blaze shape will be magnified in the divided spectrum. The ALSFS algorithm is less susceptible to this problem because of the strong constraints provided by the laboratory source, but other blaze estimation methods, including the AFS algorithm, can be strongly affected.

The AFS algorithm also has difficulty in this region in cases where the edge of an order splits an absorption line. Fortunately, neighboring orders often have some region of overlap that can be used to correct the boundaries. A weighted average of the blaze-removed spectrum of the two orders can be used as an estimate of the blaze function in the overlapping region. For example, Figure 9 shows two neighboring orders that share an overlapping region. Figure 9(a) shows the right boundary of the left order, which looks good as a blaze-removed spectrum using AFS algorithm. Figure 9(b) shows the left boundary of the right order, which spuriously rises above 1. We correct the overlapping region using the following:

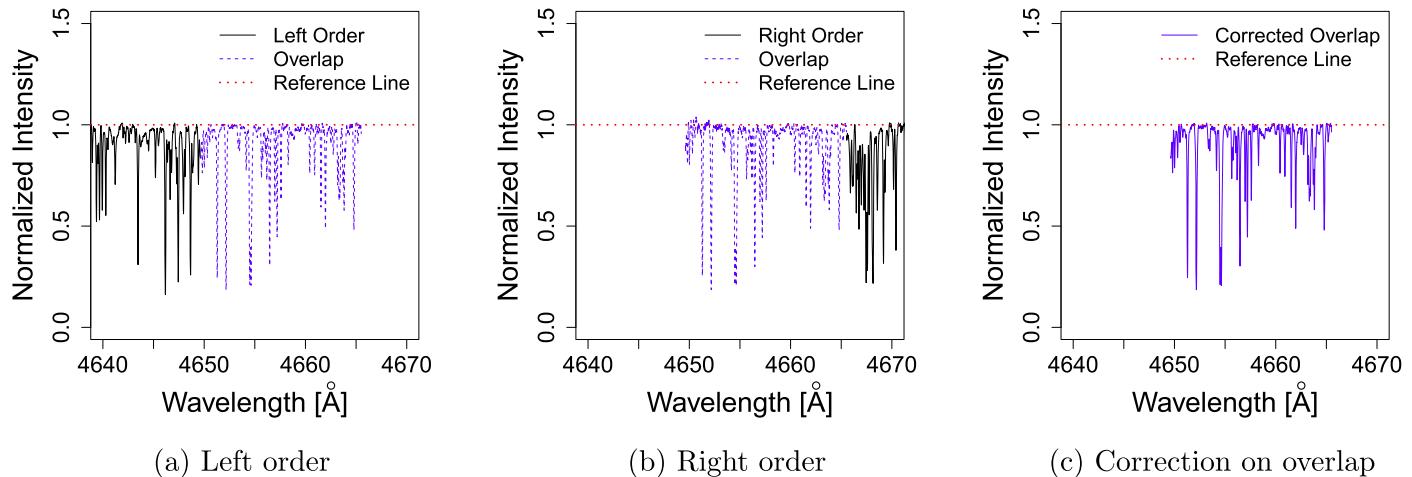
$$y_{\text{corrected},l} = w_l y_{1,l} + (1 - w_l) y_{2,l}, \quad l = 1, \dots, m,$$

where  $y_1$  and  $y_2$  are the intensities for overlapping regions from the left and right orders, respectively,  $y_{\text{corrected}}$  is the array of intensities for the corrected overlapping region,  $m$  is the number of pixels in the overlapping region, and  $w_l = 1 - \frac{l}{m}$ ,

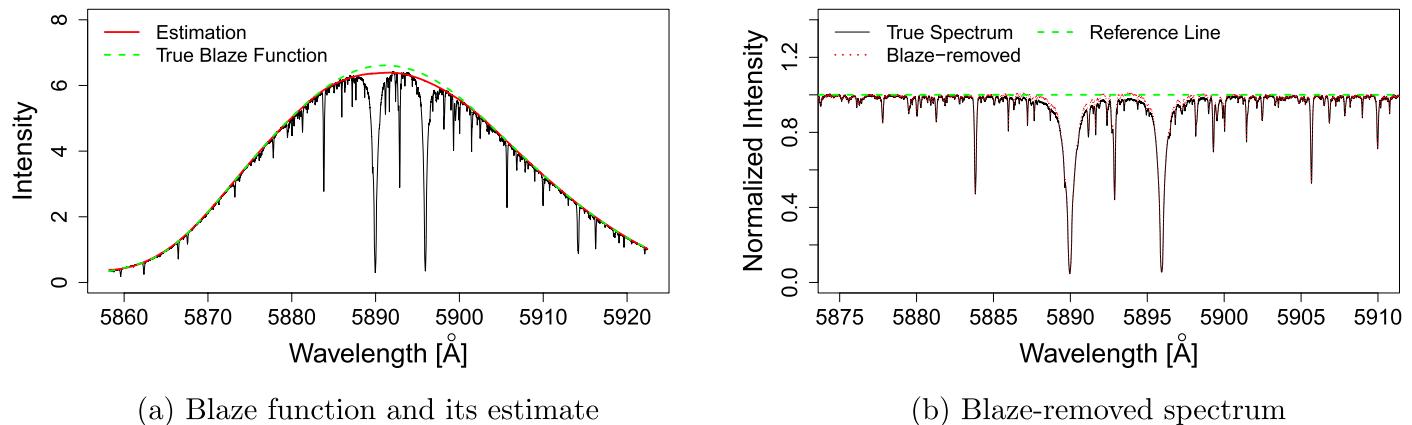
for  $l = 1, \dots, m$ . The result of the correction is shown in Figure 9(c). The boundary-corrected spectrum is much better than the original estimate, and we can achieve further improvement by changing the definition of  $w_l$ . For example, if it is known that one of the two orders has a better estimate on its boundary, we can assign more weight toward the better order. This might be the case for a pair of orders with a broad spectral feature that is cut off on only one of the orders.

#### 2.4.2. Wide Absorption Lines

Sometimes an order has wide absorption lines that influence the performance of the blaze function estimation. Wide absorption lines can significantly influence the AFS algorithm performance, but only slightly influence the ALSFS algorithm. Figure 10(b) shows the order containing the two deep Na D lines, which are each so broad that the spectrum does not fully return to the nominal continuum level between them. When attempting to fit the blaze function, the alpha shape will dip into wide features like these and pull the final blaze function estimate downward. In Figure 10(a), the spectrum (same as in Figure 10(b)) before the blaze function removal is displayed. Despite failing to return to the proper continuum level, this attempt at normalization looks reasonable to the eye. Without prior information, the proposed data-driven AFS algorithm cannot consistently determine the continuum level over regions of greatly extended absorption like this one. If we know there is a wide absorption feature before applying the algorithm, the regions can be masked in step 4 of the AFS algorithm, excluding those points from the estimate.



**Figure 9.** Combining neighboring orders to correct boundary estimations. (a) The right part of the left order. The blue dashed line is the overlap region, which is shared by the two orders. (b) The left part of the right order, with the blue dashed line again showing the overlapping region. (c) Overlapping region after correction using a weighted average of the two orders.



**Figure 10.** Wide absorption features are a problem for the AFS algorithm without prior information. (a) A spectrum with blaze function is shown in black solid line. The estimation of the AFS algorithm is shown in red solid line. The true blaze function is shown in green dashed line. (b) The true spectrum without blaze function is shown in black solid line. The spectrum after blaze function removal by the AFS algorithm is shown in red dashed line. A reference line at  $y = 1$  is shown as green dashed line.

### 2.4.3. Continuous Opacity

If an absorption region is wider than half of the wavelength range of the target order, we refer to it as *continuous opacity*. This problem impacts both the AFS algorithm and the ALSFS algorithm. Because the region is so wide, masking is not an option for the AFS algorithm. The only way to deal with it is to use prior information about continuous opacity: location and intensity. Then the spectrum can be adjusted by accounting for the opacity to get a good estimation. The ALSFS algorithm can address this by connecting and adjusting neighboring orders. Because there are overlaps between neighboring orders, one can search neighboring orders until finding an order that does not contain continuous opacity as a reference. Then the orders with continuous opacity can be adjusted to the level of the reference order. An example is shown in Section 3.4 to illustrate the continuous opacity correction.

### 3. Simulations

In our simulation study, we use an integrated-disk solar flux atlas spectrum (Wallace et al. 2011) produced by the National

Solar Observatory (NSO), obtained with the McMath-Pierce Solar Telescope's Fourier transform spectrometer. Because the spectra in the atlas were obtained with a Fourier transform spectrometer rather than an echelle spectrograph, they have no intrinsic blaze function. The atlas has been approximately continuum normalized. The spectral resolution of the atlas ranges from 350,000 to 700,000, and the spectra are essentially noiseless.

To mimic the data characteristic of EXtreme PREcision Spectrometer (EXPRES), we use the same wavelength ranges as EXPRES to divide the NSO spectrum into artificial orders. We impose a shape based on a blaze function estimated from a B-star spectrum onto each order and use our algorithms to remove the blaze function. Then simulated photon noise (Gaussian white noise, which well approximates Poisson noise for high S/N) is added corresponding to a S/N of 300. The noisy blaze-imposed spectrum is divided by the true blaze function of produce a benchmark flattened spectrum. In this simulation study, we use two orders: one from the bluer end of the spectrum and one from the redder end. The AFS algorithm and the ALSFS algorithm are tested on the two orders,

respectively. Additionally, we compare our algorithms with the commonly used *iterative method*. The iterative method is introduced next.

### 3.1. Iterative Method

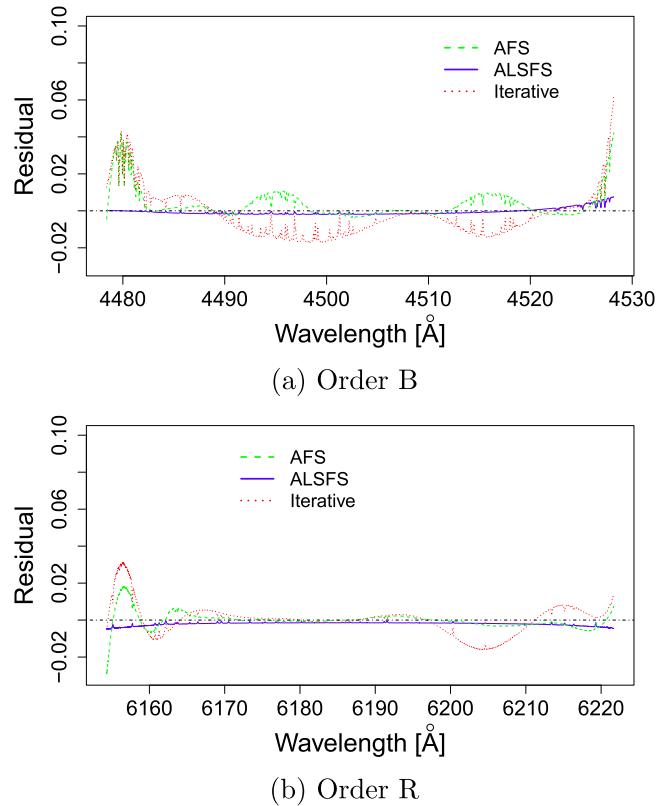
The iterative method is commonly used to remove blaze function from a spectrum. A polynomial model is fit to a spectrum order, and the fit is considered as the starting estimation for the blaze function. Next, the method iterates. In each iteration, there is a threshold curve obtained by:

$$\text{threshold}(\lambda_i, t) = \text{fit}_{\text{polynomial}}^{(t)}(\lambda_i) - \frac{0.5}{t+1}, \quad (2)$$

where  $\text{fit}_{\text{polynomial}}^{(t)}(\lambda_i)$  is the seventh-order polynomial regression estimate at  $\lambda_i$  and  $t$  is the iteration time. As  $t$  increases, the threshold curve increases. The method builds a subset  $M_t$  containing wavelength  $\lambda_i$ 's with intensity larger than  $\text{threshold}(\lambda_i, t)$ . Then in the next iteration, it only uses spectrum points whose wavelength values are in  $M_t$  to fit a polynomial model and the fit is a new estimation for continuum. To stop the iteration, a stopping time, denoted as  $T$ , can be set at a particular iteration. Another option is to set a standard deviation value (sd), such that the algorithm stops when the standard deviation of the blaze-removed spectrum is smaller than sd. The sd cannot be too small (much smaller than the standard deviation of the true spectrum without the blaze function), otherwise the iteration will not stop. In this work, we stop the iteration if either it arrives at the  $T$ th iteration or the standard deviation of the blaze-removed spectrum is smaller than sd. Empirically, an sd value around 0.05 works well, but the performance of the iterative method is influenced by the choice of  $T$ . A higher S/N requires a larger  $T$  and, in general, we have found a  $T$  value equal to S/N seems to well.

### 3.2. Simulated Spectra

The first order (blue) has wavelengths ranging from 4478 to 4528 Å, called “order B,” and the second order (red) has wavelengths ranging from 6154 to 6221 Å, called “order R.” For this simulation, we require a realistic blaze function to add to the NSO spectrum. To do this, we estimate a blaze function of a B-star, HR 5501, observed with EXPRES (Jurgenson et al. 2016). We first use the ALSFS algorithm, with the corresponding LED spectrum (after using the AFS algorithm on it) as the laboratory source, on the B-star spectrum to get the estimate of the blaze function. Then we apply this estimate as the blaze function to the two orders. The raw spectrum of the B-star spectrum is then used as the laboratory source reference for the ALSFS algorithm. The AFS algorithm, ALSFS algorithm, and iterative method are applied to estimate blaze functions of the two orders and residuals are calculated. The residuals of order B and R with S/N = 300 are displayed in Figures 11(a) and (b), respectively. Overall, ALSFS has the smallest residuals that are consistent across the whole order. AFS has larger residuals than ALSFS, which increase in the boundary regions, but the iterative method has larger residuals than AFS in both boundary regions and middle regions.



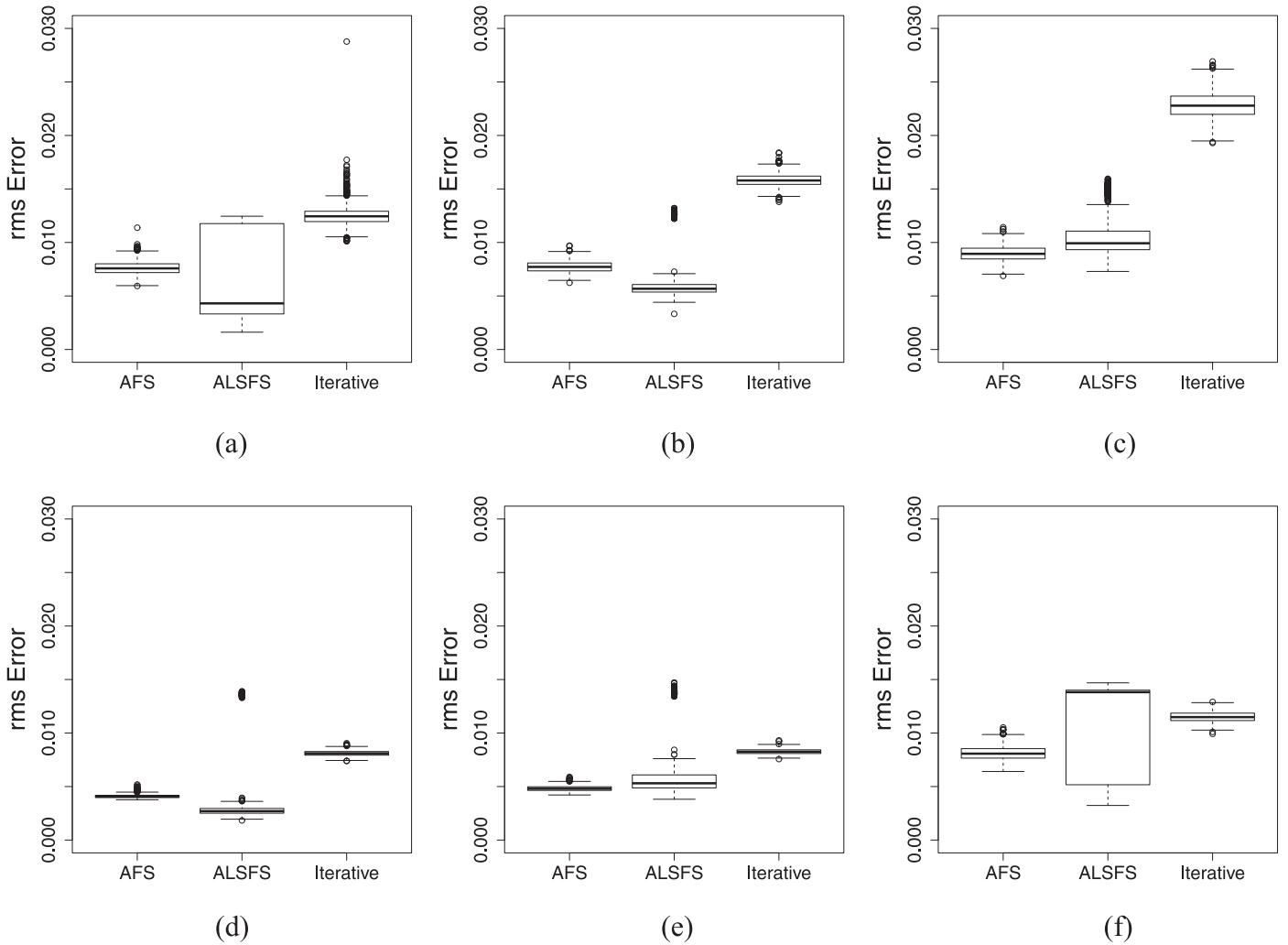
**Figure 11.** Two orders from NSO data, with wavelength range according to EXPRES spectrum. Order B is from wavelength range 4478–4528 Å and the order R is from wavelength range 6154–6221 Å. Three methods are applied to the two orders to compare residuals. (a) Residuals of order B. (b) Residuals of order R. AFS displayed here are without the boundary corrections—it could be improved further by boundary modification described in Section 2.4.1.

## 3.3. Results

Figure 11 displays a single realization of the noisy spectrum. We repeat the procedure by adding different realizations of noise 1000 times. For each of the 1000 realizations, the AFS algorithm, ALSFS algorithm, and the iterative method are applied to the two orders. This is carried out for an S/N 300, 150, and 50. The results are displayed in Figure 12 and Table 1. The rms error is calculated as  $\sqrt{\frac{1}{n} \sum_{i=1}^n r_i^2}$ , where  $r_i$  is the residual at pixel  $i$ . The ALSFS algorithm has the smallest median rms error for three scenarios: S/N 300-order B, S/N 300-order R, and S/N 150-order B. For the other three scenarios, the AFS algorithm has the smallest median rms error. Except for S/N 50-order R, the iterative method has the largest median rms error.

### 3.4. Correction for Continuous Opacity

An example of continuous opacity is illustrated using the same simulation setup: we apply a blaze function obtained from a B-star spectrum to the NSO spectrum, and then we attempt to recover the underlying spectrum. Here we examine five consecutive artificial orders. A region of continuous opacity spans about 80 Å, which is captured in parts of the third, fourth, and fifth orders, displayed in Figure 13(a). After applying the ALSFS algorithm, the resulting blaze-removed spectra are displayed in Figure 13(b). Although the blaze-removed spectra are flat, the ALSFS algorithm does not capture the continuous opacity correctly on its own. We know that the first two orders are not affected by the continuous opacity, and so they are used



**Figure 12.** Distribution of rms errors over 1000 samples for S/N 300, 150, and 50. The simulation is repeated 1000 times for the three methods on the two orders, respectively. (a)–(c) are results of order B. (d)–(f) are results of order R. Because the true spectra are known, the rms errors could be calculated.

**Table 1**

Unit:  $1 \times 10^{-3}$ . Medians of rms Errors Over 1000 Samples for Orders B and R with Varying S/N for AFS, ALSFS, and the Iterative Method

S/N	Order B			Order R		
	AFS	ALSFS	Iterative	AFS	ALSFS	Iterative
300	7.58	4.31	12.44	4.06	2.70	8.10
150	7.71	5.67	15.79	4.80	5.31	8.25
50	8.94	9.93	22.78	8.07	13.82	11.49

**Note.** Medians listed in this table are corresponding to distributions of rms error in Figure 12.

as the reference to correct the other orders: each order is linearly adjusted in intercept and slope to be align with its left neighboring order using the points within the overlapping region whose normalized intensities are in the top  $\tilde{q}$  quantile. Ideally, the combined segment should recover the continuous opacity in the normalized intensities after the blaze function removal. However, error in the slope estimation of the first order is amplified: the combined spectrum in Figure 13(c) is not perfectly horizontal and it goes upward from left to right. We fit another linear regression using the points in the combined spectrum whose normalized intensities are in the top

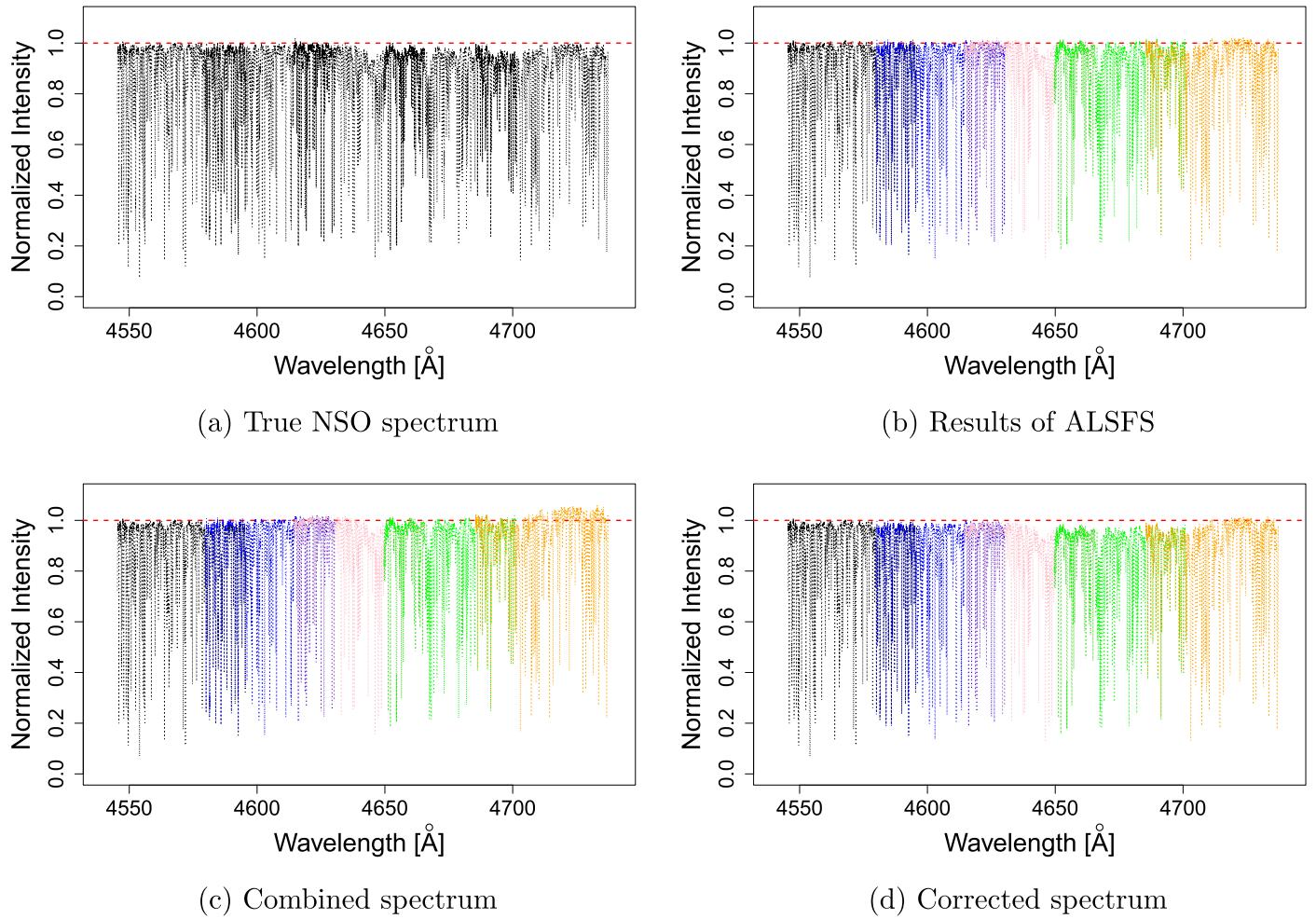
$\tilde{q}$  quantile to remove the extra slope. The resulting spectrum recovers the continuous opacity well, as displayed in Figure 13(d).

### 3.5. Cosmic Rays

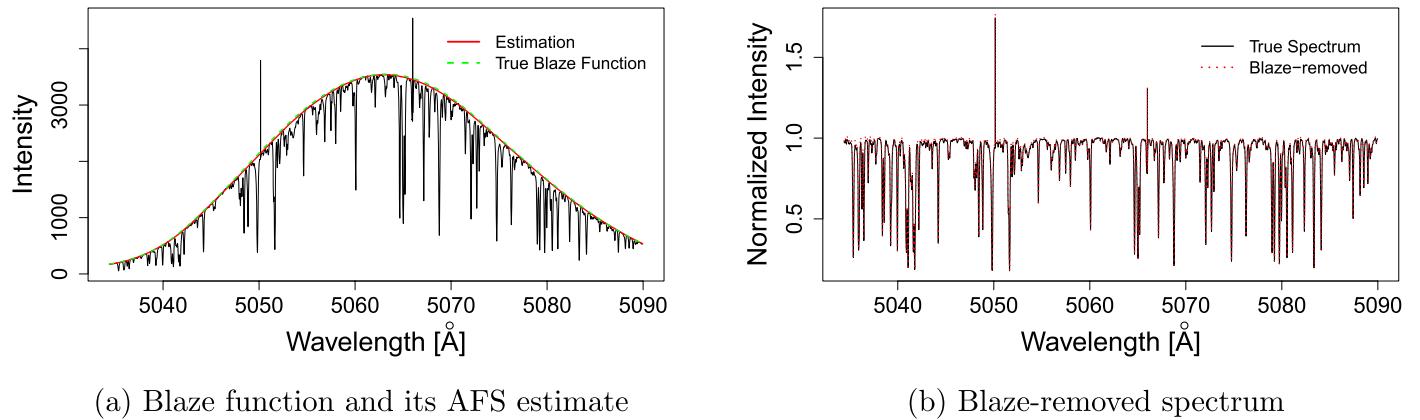
The proposed methods can also be used for spectra with cosmic rays. In particular, the modified AFS algorithm for laboratory source smoothing, described in Section 2.2.2, can be used directly to deal with the presence of cosmic rays. To demonstrate this, an order with simulated cosmic rays is displayed in Figure 14(a), which contains two upward spikes. The blaze function estimate from the AFS algorithm is shown in Figure 14(a), and the blaze-removed spectrum is displayed in Figure 14(a) in comparison to the true spectrum. The ALSFS algorithm can be modified similarly to deal with cosmic rays.

## 4. Applications and Discussions

The AFS algorithm can be useful for studying telluric and microtelluric absorption lines. Telluric lines, originating in the Earth's atmosphere, create time-varying and humidity-dependent perturbations to the shapes of stellar lines (Leet et al. 2019). It is a particularly acute problem in the field of high-precision exoplanet radial velocity detection where uncorrected telluric



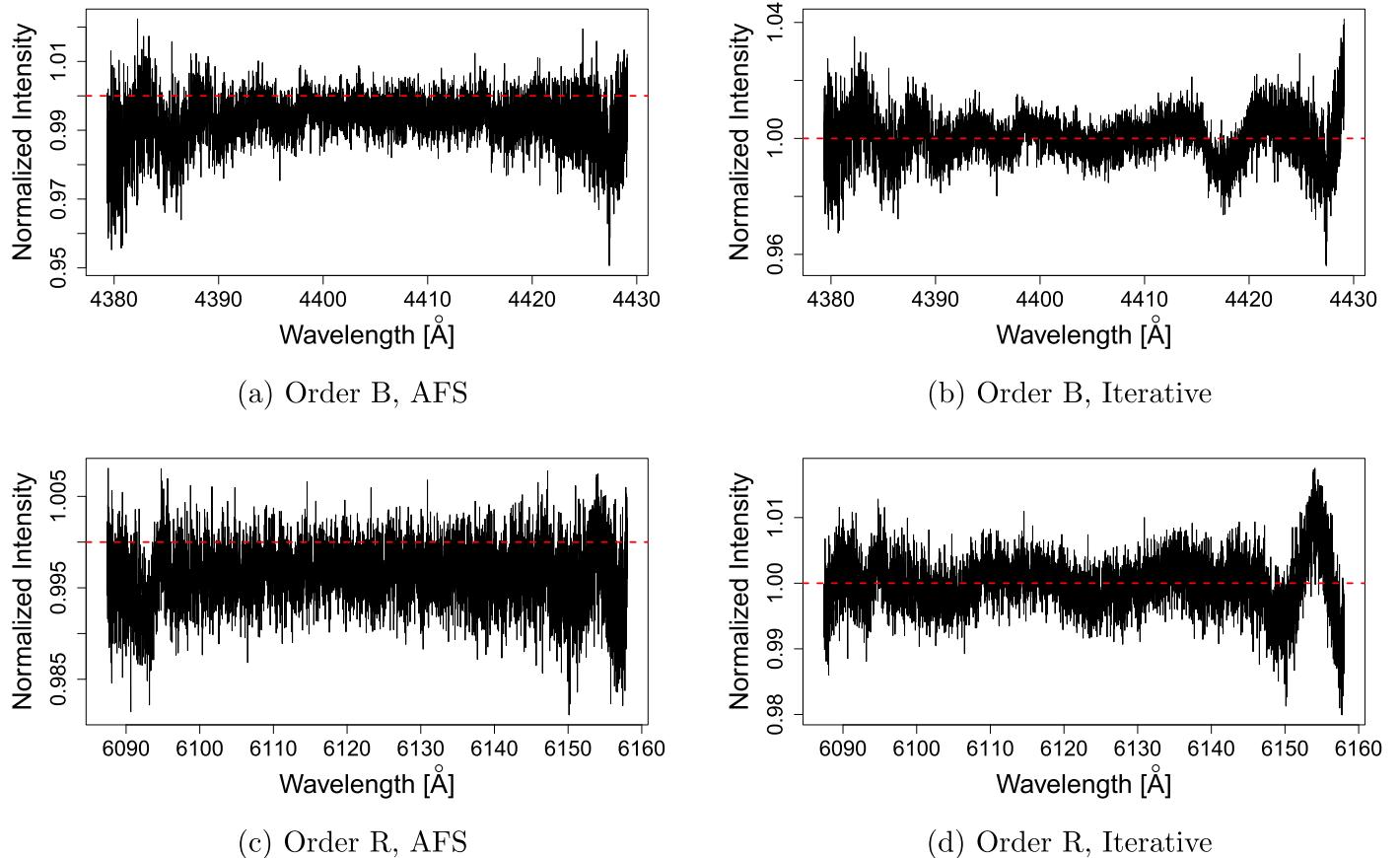
**Figure 13.** (a) Segment of the full NSO spectrum, with significant continuous opacity redwards of about  $4630 \text{ \AA}$ . The red dashed line is a reference line at normalized intensity = 1. (b) The segment is divided into five artificial orders with injected blaze functions. Here we show the resulting blaze-removed orders from ALSFS shown in different colors, respectively. (c) Each order is linearly adjusted with its blueward neighboring order as a reference based on their overlapping region. The combined long spectrum is still imperfect because the slope estimation of the first order could be imperfect. (d) An ordinary linear regression is fitted to the combined spectrum to remove the extra slope. The resulting spectrum recovers the continuous opacity well.



**Figure 14.** (a) Order with simulated cosmic rays. The black solid line is the raw spectrum, the red solid line is the blaze function estimate from the AFS algorithm, and the green dashed line is the true blaze function. (b) The blaze-removed spectrum using the AFS algorithm. The black solid line is the true spectrum without the blaze function and red dashed line is the blaze-removed spectrum.

lines contribute a radial velocity error of  $\sim 0.2$  to 1 meters per second in optical wavelengths (Cunha et al. 2014) and as much as a few meters per second in near-infrared wavelengths (Bean et al. 2010).

Telluric lines can be measured using spectroscopic observations of B-stars, which are bright, rapidly rotating young stars whose spectra are devoid of all but the strongest absorption lines because of extreme rotational broadening. A B-star acts as



**Figure 15.** B-star spectrum, HR 8634, observed with EXPRES on 2018 July 7. The AFS and iterative method are applied to a blue order and a red order respectively. Panels (a) and (c) are the blue and red order flattened by AFS. Panels (b) and (d) are the blue and red order flattened by the iterative polynomial fitting method, which leaves a zigzag shape to the continuum.

a background against which narrow telluric lines can be observed as a calibration tool for radial velocity measurements of other stars. Fitting and removing the blaze function and continuum of B-stars allows the depth of the telluric lines to be measured, which folds into current and proposed methods to mitigate their effects (e.g., Leet et al. 2019).

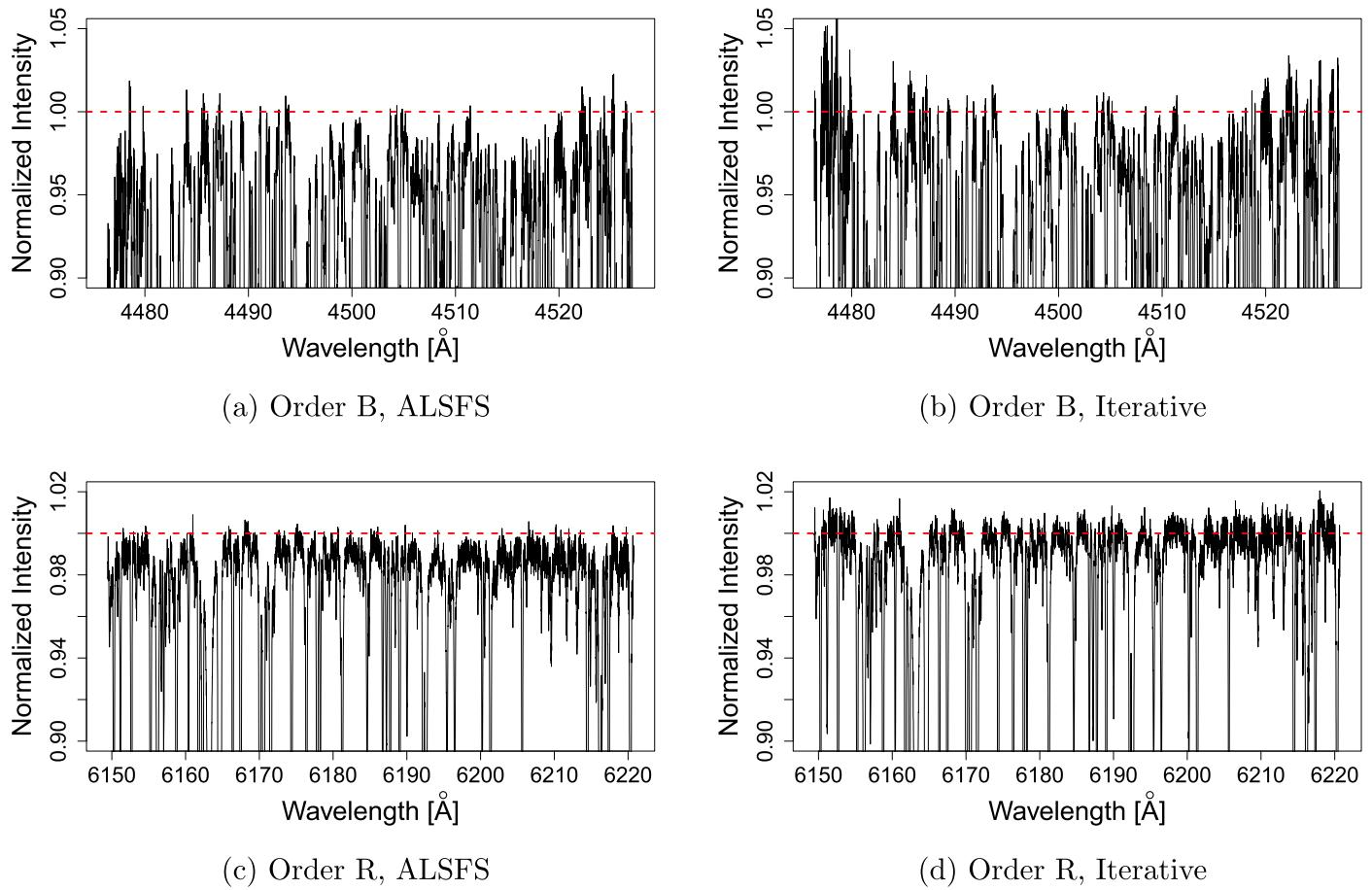
To illustrate the applicability of the proposed algorithms, the AFS and, for comparison, the iterative method are applied to a B-star spectrum, HR 8634, which was observed with EXPRES (Jurgenson et al. 2016) on 2018 July 7. A blue order (order B: 4376–4431 Å) and a red order (order R: 6085–6160 Å) of this spectrum were selected to show the effect of the blaze function removal. The flattened spectra are displayed in Figure 15, with a reference line at normalized intensity = 1. Though the ground truth is unknown, it appears that the proposed AFS works well on flattening the spectra, while the spectra flattened by the iterative method have a zigzag pattern.

The ALSFS method can be useful for estimating the blaze function of the more complex spectra of late-type stars by incorporating information from a laboratory source spectrum to obtain an initial guess. Late-type stars are the primary targets of Extreme Precision Radial Velocity planet searches, which aim to suppress radial velocity measurement errors below  $\sim 1$  meter per second. Among the greatest challenges hindering Extreme Precision Radial Velocity is the problem of stellar activity: magnetically driven motions within the stellar atmosphere lead

to time-varying features, such as spots and faculae, that create line-profile distortions that skew the measured centroids of the lines leading to imprecise radial velocity measurements (summarized in Fischer et al. 2016, Section 4.2).

Work to address the problem of stellar activity is ongoing, but one encouraging approach used by several teams is to investigate the sensitivity of individual spectral lines to activity, in order to obtain activity-free radial velocities (Davis et al. 2017; Dumusque 2018; Wise et al. 2018). These methods all utilize some sort of continuum-fitting method, because the depths of the lines must be known with precision in order to search for correlations with activity over time. By providing flatter, more uniform blaze function estimates, the ALSFS algorithm will permit more precise measurements of the individual line depths and line-profile shapes that are correlated with stellar activity.

The ALSFS and iterative method are applied to a blue order (order B: 4473–4529 Å) and a red order (order R: 6147–6223 Å) of the star 51 Pegasi, observed with EXPRES on 2018 July 8. Amplified figures of the flattened orders are displayed in Figure 16. For order B, the spectrum from ALSFS is mostly flat, while the spectrum from the iterative method has much higher intensities on boundary regions. For order R, both methods work well in flattening, but the iterative method is inaccurate in scale so that the peaks are above the reference line at Normalized Intensity = 1.



**Figure 16.** ALSFS and the iterative method are applied to a blue order and a red order of the star 51 Pegasi, observed with EXPRES on 2018 July 8. Panels (a) and (b) are the blue order flattened by ALSFS and the iterative method, respectively. Panels (c) and (d) are the red order flattened by ALSFS and the iterative method. Both (c) and (d) are pretty flat, while (d) has a scale issue that the normalized intensities are higher than expected.

We did not test the methods on cooler stars, where the continuum is poorly defined, such as M dwarfs. Because the Sun is a relatively metal-rich star, as is 51 Peg (metallicity of +0.2; Frasca et al. 2009), our methods are expected to perform at least as well on stars with lower metallicities and higher temperatures.

## 5. Conclusions

In this work, we presented two data-driven algorithms, AFS and ALSFS, for removing the blaze function from spectra obtained from echelle spectrographs. The key aspects of the algorithms are the use of alpha shapes to provide an initial guess of the blaze function’s shape, and the use of local polynomial regression to refine this guess. The two algorithms are designed for two scenarios: the AFS algorithm operates without a reference spectrum and may be applied directly to stellar spectra containing even a high number of absorption lines, while the ALSFS algorithm also incorporates additional information from a reference continuum spectrum to inform its initial guess. As an application of the AFS algorithm, a continuum laboratory source reference spectrum—such as an LED or quartz lamp spectrum—could be corrected and smoothed to be used in the ALSFS algorithm.

A simulation study was presented to illustrate the performance of the proposed algorithms compared to the commonly used iterative method for spectral normalization. In general, our

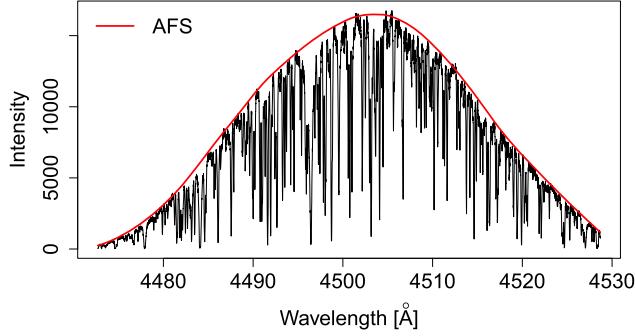
algorithms have smaller rms errors than the iterative method. Overall, the ALSFS algorithm has the smallest median rms error when S/N is high. Moreover, our algorithms are able to capture the edge effects better than the iterative approach. ALSFS is relatively robust to edge effects, and we have also developed a method of boundary correction for the AFS algorithm. Furthermore, detailed discussion regarding the applications of the algorithms was presented with examples of B-star and star 51 Pegasi spectra, which are observed with EXPRES.

This work proposes a method to correct the continuum of an echelle spectrum by modeling the blaze function of individual orders. A flattened echelle spectrum obtained from the proposed methods works better than its original form in studying physical and astronomical properties of a star, e.g., blaze-removed B-star spectra for understanding telluric lines, more precise absorption line depths for studying stellar activity.

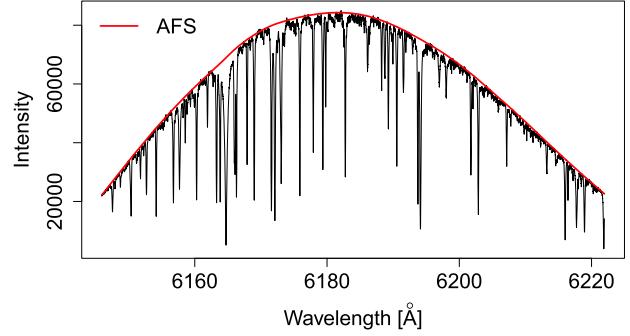
J.C.K. and D.F. gratefully acknowledge support through NSF 1616086, NASA XRP 80NSSC18K0443. A.B.D. acknowledges support through the NSF Graduate Research Fellowship grant DGE1122492.

## Appendix

The selection of the parameter  $\alpha$  is discussed in Section 2.3 so in this section we focus on the selection of  $q$  and  $m_0$ . The  $m_0$



(a) A blue order of 55 Cancri, S/N 130, temperature 5165 K, metallicity +0.27 dex:  $q = 0.99$ ,  $m_0 = 0.35$

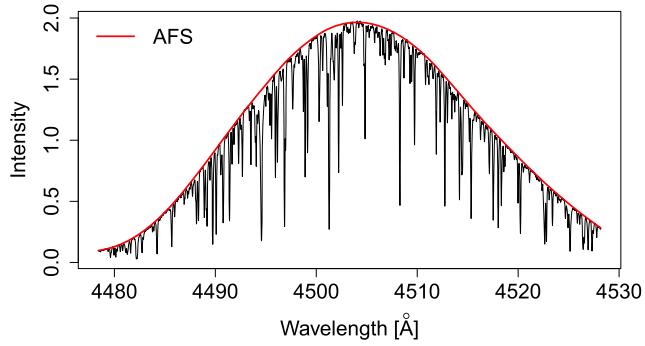


(b) A red order of 55 Cancri, S/N 290, temperature 5165 K, metallicity +0.27 dex:  $q = 0.97$ ,  $m_0 = 0.25$

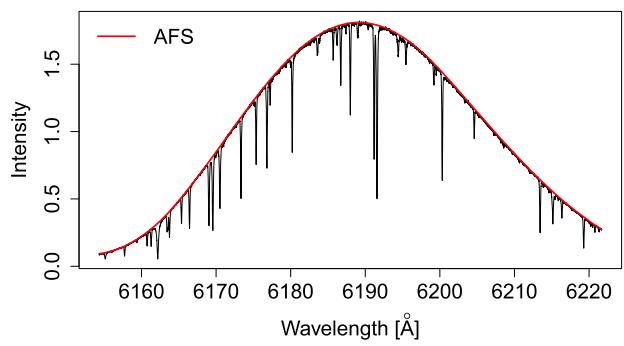
**Figure 17.** Examples for parameter selection.

depends on the amount of absorption of an order, and  $q$  depends on both the S/N and the amount of absorption. While S/N can be estimated, the amount of absorption is influenced by multiple factors such as wavelength, temperature, surface gravity, and stellar metallicity. Instead of recommending parameter values based on these factors individually, we provide several example orders to give an idea of the rough ranges of parameters to use. The examples below are all echelle spectra, and an  $\alpha = \frac{1}{6} \times$

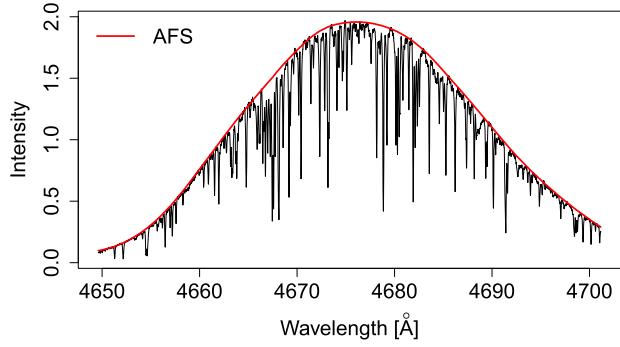
wavelength range is used for all of them. The selected  $q$  and  $m_0$  values are listed under each figure. Figure 17 displays two orders of EXPRES spectra for the G8 V star 55 Cancri, which has a temperature of about 5165 K and a metallicity of +0.27 dex (Marcy et al. 2002). Figure 18 displays six simulated orders from the NSO solar spectrum, as described in Section 3. The noise is added to each order with S/N set to be 300, 150, or 50. We suggest these examples be used as a guide for parameter selection.



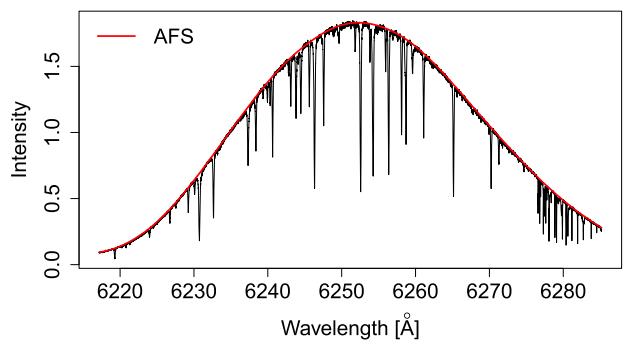
(a) A blue order of NSO simulated spectrum, S/N 300, temperature 5778 K, metallicity 0 dex:  $q = 0.99$ ,  $m_0 = 0.3$



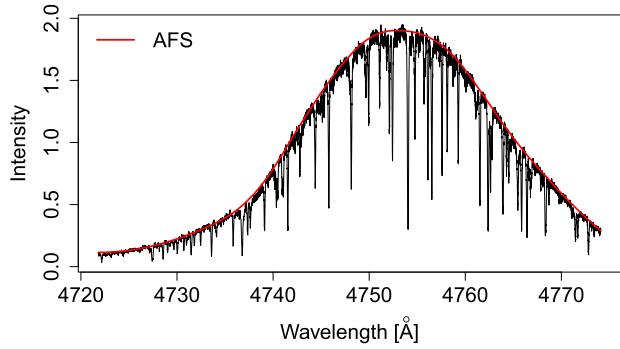
(b) A red order of NSO simulated spectrum, S/N 300, temperature 5778 K, metallicity 0 dex:  $q = 0.95$ ,  $m_0 = 0.25$



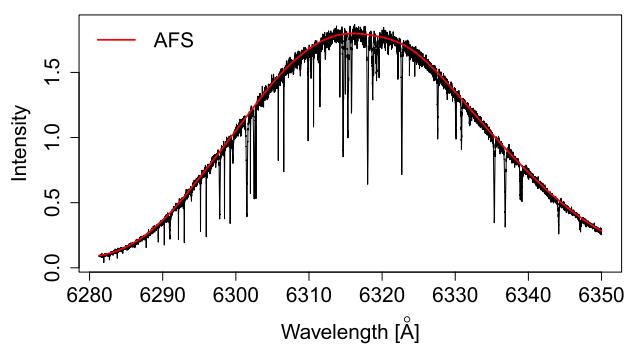
(c) A blue order of NSO simulated spectrum, S/N 150, temperature 5778 K, metallicity 0 dex:  $q = 0.95$ ,  $m_0 = 0.25$



(d) A red order of NSO simulated spectrum, S/N 150, temperature 5778 K, metallicity 0 dex:  $q = 0.8$ ,  $m_0 = 0.25$



(e) A blue order of NSO simulated spectrum, S/N 50, temperature 5778 K, metallicity 0 dex:  $q = 0.8$ ,  $m_0 = 0.25$



(f) A red order of NSO simulated spectrum, S/N 50, temperature 5778 K, metallicity 0 dex:  $q = 0.5$ ,  $m_0 = 0.25$

**Figure 18.** Examples for parameter selection.

## ORCID iDs

- Xin Xu <https://orcid.org/0000-0001-5574-745X>  
 Jessi Cisewski-Kehe <https://orcid.org/0000-0002-9656-2272>  
 Allen B. Davis <https://orcid.org/0000-0002-5070-8395>  
 Debra A. Fischer <https://orcid.org/0000-0003-2221-0861>  
 John M. Brewer <https://orcid.org/0000-0002-9873-1471>

## References

- Barker, P. K. 1984, *AJ*, **89**, 899  
 Bean, J. L., Seifahrt, A., Hartman, H., et al. 2010, *ApJ*, **713**, 410  
 Blanco-Cuaresma, S., Soubiran, C., Jofré, P., & Heiter, U. 2014, *A&A*, **566**, A98  
 Cleveland, W. S. 1979, *J. Am. Stat. Assoc.*, **74**, 829  
 Cunha, D., Santos, N., Figueira, P., et al. 2014, *A&A*, **568**, A35  
 Davis, A. B., Cisewski, J., Dumusque, X., Fischer, D. A., & Ford, E. B. 2017, *ApJ*, **846**, 59  
 Dumusque, X. 2018, *A&A*, **620**, A47  
 Edelsbrunner, H., Kirkpatrick, D., & Seidel, R. 1983, *ITIT*, **29**, 551  
 Fischer, D. A., Anglada-Escude, G., Arriagada, P., et al. 2016, *PASP*, **128**, 066001  
 Frasca, A., Covino, E., Spezzi, L., et al. 2009, *A&A*, **508**, 1313  
 Jurgenson, C., Fischer, D., McCracken, T., et al. 2016, *Proc. SPIE*, **9908**, 99086T  
 Leet, C., Fischer, D. A., & Valenti, J. A. 2019, *AJ*, **157**, 187  
 Marcy, G. W., Butler, R. P., Fischer, D. A., et al. 2002, *ApJ*, **581**, 1375  
 Skoda, P., Surlan, B., & Tomic, S. 2008, *Proc. SPIE*, **7014**, 70145X1  
 Torres, G., Fischer, D. A., Sozzetti, A., et al. 2012, *ApJ*, **757**, 161  
 Wallace, L., Hinkle, K., Livingston, W., & Davis, S. 2011, *ApJS*, **195**, 6  
 Wise, A., Dodson-Robinson, S., Bevenour, K., & Provini, A. 2018, *AJ*, **156**, 180