# Automatic Normalization and Equivalent-Width Measurement of High-Resolution Stellar Spectra *

Jing-Kun Zhao, Gang Zhao, Yu-Qin Chen, Jian-Rong Shi, Yu-Juan Liu and Ju-Yong Zhang

National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012; *zjk@yac.bao.ac.cn*

**Abstract** It is well known that normalization, radial velocity correction and equivalent-width measurement of high-resolution stellar spectra are time-consuming work. In order to improve the efficiency we present an automatic method for these routines. The continuum is determined by fitting the 'high points' in the spectrum. After continuum normalization, the program automatically searches for the position of the H$\alpha$ line and obtains a rough radial velocity, then computes an accurate radial velocity by cross-correlation between the given spectrum and the solar spectrum. In this method, the equivalent-width is automatically measured using Gaussian fitting. A comparison between our results and those from traditional analysis shows that the typical error for equivalent width is around 3.8% in our method. Developing such automatic routines does not mean to replace the interactive reduction method: it is just for a quick extraction of information from the spectra, especially those obtained in large sky surveys.

**Key words:** techniques: radial velocities – stars: abundances

## 1 INTRODUCTION

Elemental abundances are usually determined by high-resolution spectroscopy. MIDAS is one of the tools for spectroscopic data reduction, in which the correction for Doppler shift, continuum normalization and equivalent-width measurement are carried out interactively. This means it takes a rather long time to complete these reductions.

Defining the continuum, even for spectra observed without any distortion, is a complicated task. At present, when using MIDAS, one needs to select a number of pre-defined continuum windows to fit a low order spline, but the fitting is not always satisfactory. We need to interactively select some points to add or to delete in some regions where are no strong lines, and then fit again. Normally, this procedure has to be repeated several times before the final solution is derived. So, an automated routine is necessary. Since the continuum is in low-frequency region, we have tried to take the Fourier transform to analyze the spectra and to use a low-pass filter to filter out the high-frequency components. Then, we reverse-transform the Fourier spectrum back into wavelength space. However, this method has two drawbacks: one is that it is difficult to obtain the threshold of high-frequency; the other is that the broad spectral lines will contribute to the low-frequency component (Bailer-Jones et al. 1998). Wavelet method (Daubechies 1988; Chui 1992; Ruskai et al. 1992) is a tool of time-frequency analysis, which is better than Fourier transform. It can decompose the spectrum into several scales. The original spectrum can be expressed as the sum of all the wavelet scales and a smoothed array. Then the smoothed array is considered as continuum of the original spectrum. Starck et al. (1997) have used a wavelet iterative method to determine the local continuum and obtained good result. However, when automatic iteration is employed, problems will arise from regions of strong lines. Luo et

---

al. (2001) developed an automatic pipeline for estimating the continuum level using Daubechies wavelets. However, it is mainly designed for galactic spectra. For the wavelet method the continuum estimation involves a convolution of the spectrum by a very wide filter, and the presence of strong emission or absorption bands may cause problems as well (Starck et al. 1997). Since our high-resolution spectrum has many strong absorption lines, it is not desirable to determine the continuum by the wavelet method. Although Starck et al. (1997) obtained good result by extracting strong band in advance, this method cannot acquire the continuum in a fully automatic way. Bailer-Jones et al. (1998) used a combination of median and boxcar filtering of a spectrum to obtain its continuum, but the method requires knowing the position of strong line in advance, so it is difficult to make an automatic process.

At present, there are mainly two methods for measuring radial velocity: one is to select several lines to measure their wavelength, and then to calculate the radial velocity line by line and then derive the final radial velocity by averaging; another is the cross-correlation technique, in which we compute the cross-correlation function between the observed spectrum and the template spectrum, and then the peak of the function corresponds to the radial velocity. The template spectrum is mainly set up in two ways. One is the synthetic spectrum computed by the Kurucz Model (Kurucz 1993; Skuljan et al. 2000), another is by using the spectrum of standard star.

Generally, we use a number of lines to determine the abundance. First we set up a line table which includes the elements we need, and then measure the equivalent widths for the spectral lines in the table. For each spectral line we should interactively set two control points for measuring the equivalent width, and it usually takes a quiet long time to calculate the equivalent widths of stars. With the anticipated flood of spectroscopic data being processed, automatic measurement is indispensable. Patchett et al. (1973) have developed a program to measure the equivalent width automatically. In their program, the line profile was approximated by a triangle. The accuracy is adequate for the weak lines with no noticeable wings. It is, however, inadequate for strong lines.

In Section 2, we introduce in detail our automatic method for deriving the continuum. Our method of obtaining the radial velocity is presented in Section 3. The equivalent-width measurement is illustrated in Section 4. Finally, an application of our automatic method in analyzing the high-resolution spectra and a comparison between our results and those obtained interactively with MIDAS is given in Section 5.

## 2 AUTOMATIC METHOD OF OBTAINING CONTINUUM

Our program can automatically search 'high points' in the spectrum, then the continuum is obtained by fitting these points using a low order polynomial. The continuum is determined only from the original spectrum, and it does not need any prior assumption concerning the location of the absorption line.

### 2.1 Deletion of the Strong Lines

At the outset, the strong lines should be deleted. Now, what we have is just the spectrum and it is difficult to directly identify the strong lines. If we draw a spectrum on the computer screen, however, we can easily recognize them by the eye. Drawing spectrum on the screen is a mapping of flux values to screen pixel values. On the other hand, when there is no noise, the profile of a strong line can be described by three control points: left, right and center. Comparing the difference of pixel values between the left/right control point and the central point, the strong lines can be found. Imitating these procedures, we first transform the data with

$$S'(i) = \frac{(\max - S(i)) \times 800}{\max}, \tag{1}$$

where $S(i)$ is the flux at the point $i$, max represents the maximum flux of the spectrum, 800 is the approximate pixel number of the screen in the vertical direction; $S'(i)$ is the pixel value on screen at the point i. Secondly, we compute all the inflexions. Normally a strong line has three inflexions, while the continuum has no inflexion. However, because of noise there may be some inflexions in the continuum. The strong lines can be distinguished from noise by the line depth. For each inflexion, we compute its differences from the preceding and the following point. If either of the differences exceeds five pixels, then a strong line is indicated around this region. So the points between the inflexions (including the inflexions) are not recorded. Through such procedure the most strong lines will be removed (see the upper right panel of Fig. 1).
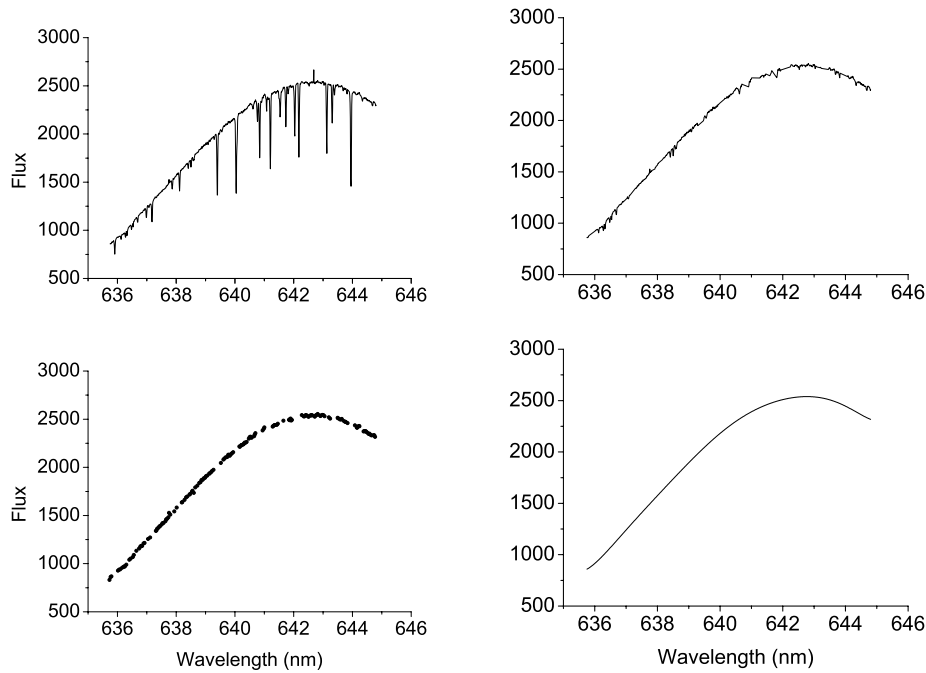
**Fig. 1** An example of the procedure for obtaining the continuum. Upper left: the original spectrum; upper right: strong lines are removed; lower left: 'high points' are determined; lower right: the continuum.

## 2.2 Recording the 'High Points'

After removing the strong lines, the program searches for the 'high points' in the remaining points by the local maximum value method. The algorithm is as follows. We define the left and right derivatives at point $i$ by

$$s'_+(i) = s(i) - s(i-1), \tag{2}$$
$$s'_-(i) = s(i+1) - s(i). \tag{3}$$

If $(s'_-(i) < 0)$ and $(s'_+(i) > 0)$ we regard the point $i$ as a local maximum point, namely a 'high point'. Going through all the spare points, all the 'high points' are obtained (see the lower left panel of Fig. 1).

## 2.3 Polynomial Fitting

Now, the spare points are the 'high points' which need to be checked before fitting. If the interval between two adjacent points is too large, we have to insert some points, otherwise the fitting result cannot accurately represent the continuum. Therefore, a maximum interval value Vc between the two adjacent points is set and Vc= $30 \times \Delta\lambda$ ($\Delta\lambda$ is the step of the original spectrum) is taken. Then the interval value between every two adjacent points is calculated. If the value>Vc, we reload the original spectrum between these two points. New 'high points' between these two points will be produced after repeating the above procedure (Sect. 2.1–2.2). At last, we fit the final points with a low order polynomial and the result is the continuum (see the lower right panel of Fig. 1). After the original spectrum is divided by the continuum, we obtain the normalized spectrum. As shown in Figure 2, the continuum and normalized spectrum derived using our program are rather satisfactory.
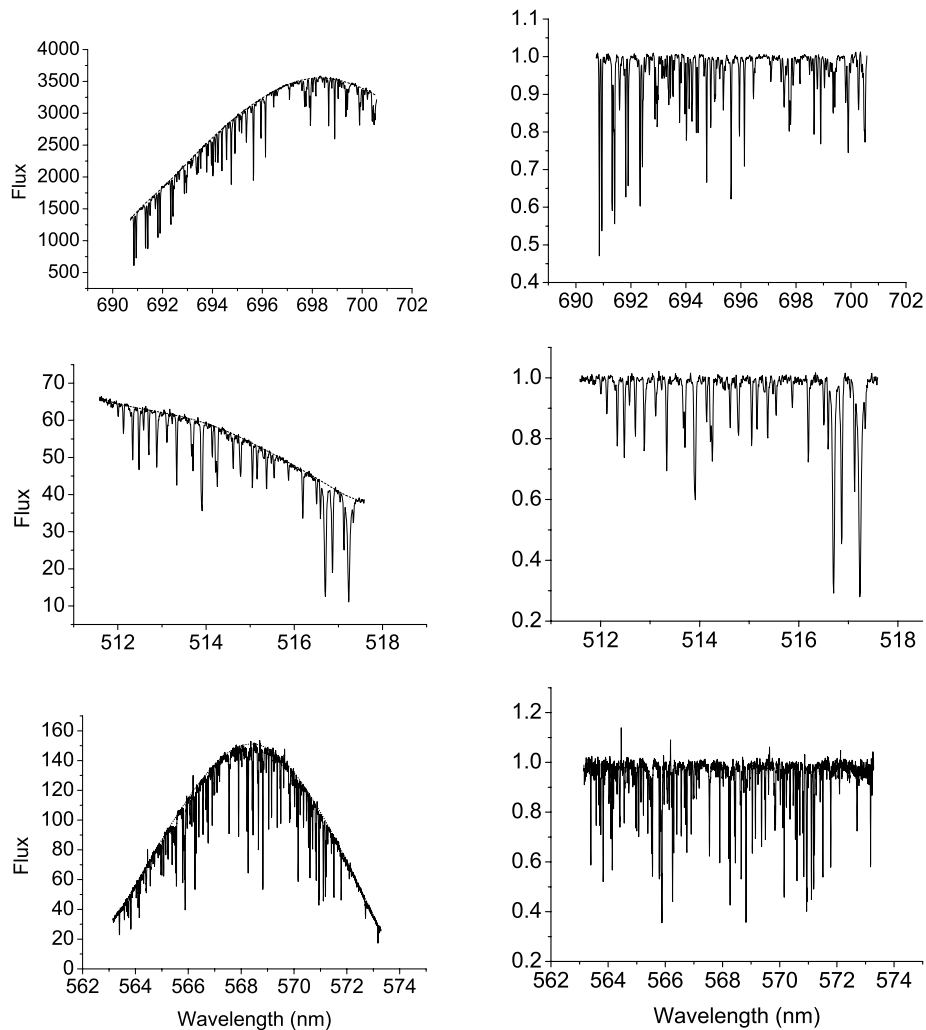
**Fig. 2** Three examples using our method to obtain normalized spectra. The left panels show the original spectra plotted with solid line and the continuum plotted with dot line. The right panels show the normalized spectra. Upper: the spectrum observed with the red path at the 2.16 m telescope in Xinglong, China; middle: the spectrum observed with the blue path at the 2.16 m telescope in Xinglong, China; bottom: the spectrum observed using the 1.88m telescope in Okayama, Japan.

## 3  THE AUTOMATIC CORRECTION OF THE DOPPLER SHIFT

Our program can automatically search the wavelength of the Hα line in the observed spectrum, then the difference between this wavelength and the wavelength of Hα is regarded as rough radial velocity correction. Based on these preliminary results, the accurate radial velocity is obtained using the cross-correlation technique.

### 3.1  The Rough Correction by Hα Line

To make certain the position of Hα line in the spectrum, we use a window covering the wavelength range 6510–6610 Å, then calculate the convolution between the spectrum $s(\lambda)$ and a Gaussian profile $F(\lambda)=a_1 \times e^{-a_2 \times \lambda^2}$. Here, $a_1$ is a measure of the depth of the profile, $a_2$, one of the width. By trial and error we found

that, when $a_1$=0.8, $a_2 = -0.15$, F can approximately represent the profile of the H$\alpha$ line. The convolution is calculated by Equation (4),

$$\sigma(k) = \sum_{\lambda} (s(\lambda) * F(\lambda - k)), \qquad 6510\,\text{Å} < k < 6610\,\text{Å}, \tag{4}$$

Finally, we seek $k_0$ where $\sigma(k_0)$ is the maximum. Then $k_0$ does represent the position of H$\alpha$ line in the spectrum. The difference between the $k_0$ value and 6562.8 Å is taken as a rough estimate of the Doppler shift defining a rough estimate of the radial velocity of the star.

### 3.2 The Accurate Radial Velocity by the Cross-Correlation Method with Solar Spectrum

After obtaining the rough radial velocity, we calculate the cross-correlation function between the spectrum and solar spectrum. Let the rough radial velocity be $v_0$. Then we compute the cross-correlation by Equation (5),

$$\sigma(v) = \sum_{\lambda} (S(\lambda \times (1 + \frac{v}{c})) * Y(\lambda)), \qquad v_0 - 5 < v < v_0 + 5, \tag{5}$$

where $S(\lambda)$ stands for the observed spectrum, $Y(\lambda)$ the solar spectrum; c is the velocity of light and $v_1$ the accurate radial velocity at which $\sigma(v_1)$ is a maximum.

### 3.3 The Characteristic of Our Method

In our method, we calculate a rough shift value using the H$\alpha$ line, then obtain the cross-correlation function between the spectrum and the solar spectrum, thereby greatly reducing the computing time. At present, our method looks suitable for solar-type stars.
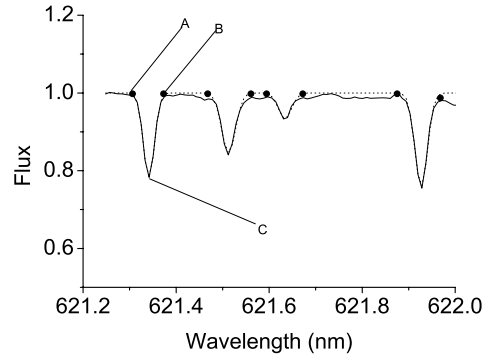


**Fig. 3** An example of automatic measurement of equivalent width. The solid line is the original spectrum, the dotted line is the Gaussian fit and dots are the control points of the spectral lines.

## 4 AUTOMATIC METHOD OF EQUIVALENT-WIDTH MEASUREMENT

The key of automatic measurement of equivalent width is to find the control points of the given spectral line. In Figure 3, 'A' is the left control point, 'B' the right control point, and 'C' the middle control point located at the central wavelength of the spectral line. In general, the control points are the inflexion points in the spectrum.

To reduce the influence of noise, we construct a new, smoothed spectrum (flux values $t(i)$) from the original spectrum (flux values $s(i)$) according to Equation (6) ($n$ the number of points in the original spectrum). At the start, the program identifies all the inflexion points of the new spectrum and saves them in

an array $P$, then it reads the line table and obtains the central wavelength of each spectral line. To each line, we query in array $P$ and get the left and the right control points which are the nearest inflexions to the central wavelength. Next the points between the two control points in the original spectrum are fitted by a Gaussian. Finally, we measure the equivalent width using the Gaussian fit. The quality of the fit can be obtained by Equations (7)–(9).

$$t(i) = \begin{cases} s(i), & i = 0,\, 1,\, n-2,\, n-1 \\ \frac{s(i-2)+s(i-1)\times 2+s(i)\times 4+s(i+1)\times 2+s(i+2)}{10}, & i = 2,\, 3\ldots n-3 \end{cases} \tag{6}$$

$$E = \frac{\sum_{i=0}^{n-1}(r(i)*g(i))}{n}, \tag{7}$$

$$\sigma = \frac{\sum_{i=0}^{n-1}((r(i)-g(i))^2}{n}, \tag{8}$$

$$L(\sigma, E) = \begin{cases} \text{`A'} & (\sigma < 0.0003 \quad \text{and} \quad E > 0.6) \\ \text{`B'} & (\sigma < 0.001 \quad \text{and} \quad E > 0.4) \\ \text{`C'} & (\sigma < 0.005 \quad \text{and} \quad E > 0.4) \\ \text{`D'} & (\text{otherwise}) \end{cases}. \tag{9}$$

Here, $r(i)$ stands for the original spectral line and $g(i)$ the fitting curve. The fitting quality is divided into four levels. To improve the precision in the determination of abundance, we can select quality 'A' or 'B' spectral lines.

## 5  A COMPARISON BETWEEN OUR AUTOMATIC METHOD AND THE INTERACTIVE METHOD WITH MIDAS

We used our automatic program to process a set of spectra observed with the Coudè Echelle Spectrograph mounted on the 2.16 m telescope of the National Astronomical Observatories (Xinglong, China). The spectra were imaged with a 1024×1024 Tektronix CCD (Zhao & Li 2001). From 2D spectra to 1D spectra, we apply the ESO MIDAS standard routines for order identification, background subtraction, flat-field correction, order extraction and wavelength calibration. The normalization, Doppler shift correction and equivalent-width measurement were done using our automatic program. Three stars, HD222368, HD9826 and HD1461, were selected to illustrate the performance of our method. The spectra of HD222368 have high S/N; those of HD9826, medium S/N; those of HD1461, low S/N. Comparisons between our results and those obtained with MIDAS are shown in Figures 4–6. For HD222368, a linear least-squares fitting with 328 common lines gives

$$\text{EW(Aut)} = (1.002 \pm 0.006)\text{EW(Int)} + 1.243(\pm 0.425), \tag{10}$$

with stand deviation 4.1 mÅ. For HD9826, a linear least-squares fitting with 310 common lines gives

$$\text{EW(Aut)} = (1.006 \pm 0.006)\text{EW(Int)} + 0.645(\pm 0.472), \tag{11}$$

with stand deviation 3.6 mÅ. For HD1461, a linear least-squares fitting with 309 common lines gives

$$\text{EW(Aut)} = (1.014 \pm 0.006)\text{EW(Int)} - 0.542(\pm 0.624), \tag{12}$$

with stand deviation 4.6 mÅ. From the above comparisons, we estimate that the average error of equivalent-width measurements related to the interactive method is about 3.8% .
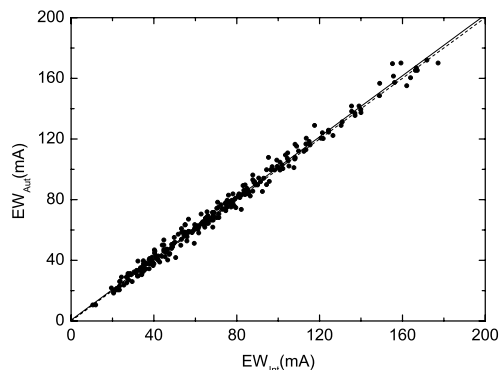
**Fig. 4** Comparison our equivalent widths, EW(Aut), with those obtained using the interactive method, EW(Int), for HD9826. The solid line is the linear least-squares fitting to the points, and the dash line is the line of equality.
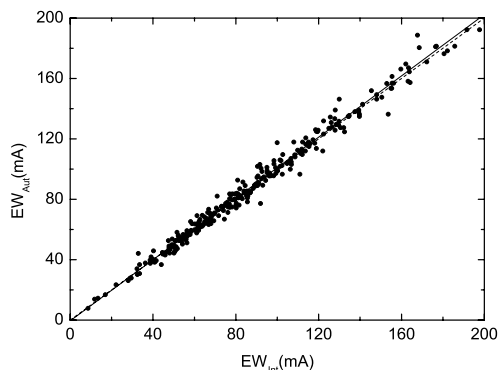
**Fig. 5** Comparison of our equivalent widths, EW(Aut), with those obtained by the interactive method, EW(Int), for HD1461. The solid line is the linear least-squares fitting to the points, and the dash line is the line of equality.
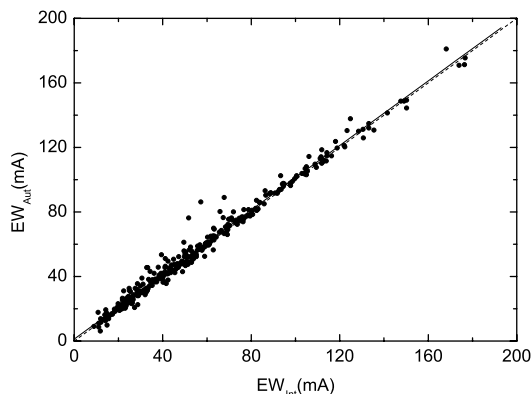


**Fig. 6** Comparison of our equivalent widths, EW(Aut), with those obtained by the interactive method, EW(Int), for HD222368. The solid line is the linear least-squares fitting to the points, and the dash line is the line of equality.

## 6 DISCUSSION AND CONCLUDING REMARKS

In our method, the continuum is determined with the original spectrum and it does not need any interactive input and both the radial velocity and equivalent-width measurements are done automatically. The efficiency of extracing information from the spectra is greatly enhanced by using our program. Table 1 compares the times taken by our method and by the interactive method, and shows that our method takes considerably less time, so demonstrating it to be a rapid way of extracting information from the spectra.

To estimate the error of the abundance calculation, we compare, for the above three stars, the calculated Fe abundances based on EW(Aut) and EW(Int). The abundance analysis is based the constant flux, homogeneous and LTE atmosphere model of Kurucz (1993). For HD222368 and HD9826, the values of $T_{\rm eff}$, $\log g$, [Fe/H] are taken from Edvardsson et al. (1993). For HD1461, the values are taken from Spinrad

**Table 1** Comparison of the times consumed between our automatic reduction and the interactive method when processing a high-resolution star spectrum, with 30 orders and 300 common lines.

|  | Radial Velocity | Normalization | E. W. Measurement |
|---|---|---|---|
| Interactive Reduction | 10 minutes | 1 hour | 8 hours |
| Automatic Reduction | 2 seconds | 5 minutes | 5 minutes |

**Table 2** The atmosphere parameters of three stars and the Fe abundance based on EW(Aut) and EW(Int)

| Star | $T_{\rm eff}$ (K) | $\log g$ | [Fe/H] (dex) | $A_{\rm EW(Aut)}$ (Fe) (dex) | $A_{\rm EW(Int)}$ (Fe) (dex) |
|---|---|---|---|---|---|
| HD1461 | 5929 | 4.0 | 0.43 | 7.61 | 7.61 |
| HD9826 | 6212 | 4.17 | 0.09 | 7.52 | 7.53 |
| HD222368 | 6255 | 4.16 | −0.17 | 7.24 | 7.23 |

et al. (1970). For all the three stars, we adopt an average microturbulence velocity of 1.5 km s$^{-1}$. Table 2 shows that the average difference in the Fe abundance for the three stars is 0.01 dex. Thus, determination of the Fe abundance using EW(Aut) is efficient.

At present our samples are mainly composed of F, G and K stars and the rotation velocities of these stars are not very large. For stars with large $v \sin i$, the normalization method will be studied in our future work. When we deal with a chromospheric active star, as only the line core of H$\alpha$ is filled in by chromospheric emission, the determination of a rough radial velocity with the H$\alpha$ line is still available, but when the H$\alpha$ line shows in emission, then our program will directly calculate an accurate radial velocity by the cross-correlation technique.

We note that one weak point of our method is that the accuracy may be affected by the quality of the spectrum. When the raw spectra are of low quality, the calculation based on our program is likely to have a low precision for individual stars, but we still can rapidly obtain statistical information for a large sample.

## References

Skuljan J., Hearnshaw J. B., Cottrell P. L., 2000, PASP, 112, 966

Bailer-Jones C. A. L., Irwin M., von Hippel T., 1998, MNRAS, 298, 1061

Luo A. L., Zhao Y. H., 2001, Chin. J. Astron. Astrophys. (ChJAA), 1, 563

Starck J. L., Siebenmorgen R., Gredel R., 1997, ApJ, 482, 1011

Patchett B. E., McCall A., Stickland D. J., 1973, MNRAS, 164, 329

Zhao G., Li H. B., 2001, Chin. J. Astron. Astrophys. (ChJAA), 1, 555

Kurucz R., 1993, Atlas Stellar Atmosphere Program and 2km/s Grid CDROM Vol. 13, Smithsonian Astrophysical Observatory, Cambridge

Edvardsson B., Andersen J., Gustafsson B. et al., 1993, A&A, 275, 101

Spinrad H., Luebke W. R., 1970, ApJ, 160, 1141

Daubechies I., 1988, Pure Appl. Math., 41, 909

Chui C. H., 1992, Wavelet Analysis and its Application, New York: Academic

Ruskai M. B., Beylkin G., Coifman R. et al., 1992, Wavelets and Their Applications, Boston: Jones Bartlett