Technical Note

Computing a Continuum and Equivalent Width for a Stellar Spectrum

by Stan Gorodenski

Equivalent Width (EW) is an important statistic for studying spectral line evolution. A good estimate of EW requires a good estimate of the continuum. Some methods assume a straight line or a visual fit of a curvilinear line as the continuum. The primary purpose of this article is to describe an alternative method for computing a continuum and EW. My method is based on least squares models and takes more time than these simpler methods, but once the computer program is written the procedure goes fairly quickly for subsequent observations of the same star, or new stars.

Although existing Windows-based spectrum processing programs can be very easy and quick to use, some have a number of disadvantages, at least from my perspective. For one, they are largely a black box, although with some study and inquiry one can deduce what is happening. Another disadvantage is that the continuum and EW computational methods are limited to what are in the programs, obviously. Also, some of the methods, as already mentioned, consist of visually fitting a line, or assuming a straight horizontal line, as the continuum to real data without statistical significance considerations. The final limitation, and to me an important one, is that the point and click menu driven methods in some of these packages make it difficult to get a record of what was done so that the results can be reproduced. The method I will describe gets around all these limitations. Also, because I did it myself I know and understand what I did, and I have the personal satisfaction of having 'built' it myself, so to speak.

A secondary purpose of this article is to show that the processing and statistical analyses of a spectrum are not mysterious and complex but simple concepts and procedures that can be performed by oneself. Although existing spectrum software packages make processing a spectrum quick and easy (I use VSPEC, IRIS, and AUDACE) they can be supplemented, as described here, when the need arises.

<u>Methods Description</u>

The spectrum has to be minimally processed to the point where it can be submitted to the scientific community for study by other researchers. Beyond this minimal processing, telluric lines have to be removed. A heliocentric correction has to be applied for line center estimation, but it is not required for continuum or EW estimation. The spectrum does not have to be normalized for the method described in this article, although the protocol is to normalize it for the scientific community.

The procedure to be described relies on two software packages. VSPEC is used to display the profile (i.e., the graph of intensity vs. wavelength), and to write it out as a .txt file for input into an SPSS (Statistical Package for the Social Sciences) computer program. SPSS models the continuum and computes EW's (as well as setting confidence limits), and is also used to produce graphs. I am biased toward SPSS because I used it as a statistical and programming package for over 20 years before I retired and so I am very knowledgeable of its programming syntax and capabilities. I was able to purchase my own stand alone copy at a reasonable price for my home PC, but other statistical packages with programming capabilities can be used, including EXCEL.

Getting a good continuum estimate for computing EW is a challenge. As already mentioned, some methods assume the continuum is a straight horizontal line. One problem with this method is that even if it is a horizontal straight line, it becomes somewhat subjective where the line should be placed vertically on the spectrum. Another problem is that not all continuums will be horizontal straight lines. I read one professional paper that proposed what appeared to be arbitrarily constructing a straight line between the beginning and end points of a spectral line. Some of the examples in the paper deviated significantly from being horizontal. The advantage of the method I will describe is that the continuum and its placement is not arbitrary but is based on least squares statistical estimates.

The method of computing the continuum in this article is an alternative to the one that uses EXCEL instead of SPSS that I described in

http://users.commspeed.net/stanlep/EquivalentWidth.doc

I developed this method and presented it at a spectrophotometry meeting held in my home in September, 2008. In that presentation I discussed what a continuum and EW are. Rather than repeating this here, the reader can read the first six pages of the referenced document document. This presentation was also reviewed by some very knowledgeable members of a discussion group called Spectro-L. Except for minor typographical errors, none of the concepts or techniques used were criticized. In a sense, it passed a peer review.

The method described in EquivalentWidth.doc for deriving a continuum is better than assuming a straight line, but it is based on visually constructing a smoothed line with point and click and a sliding bar. This curvilinear line becomes the continuum. Because it relies on a point and click and sliding bar, it is difficult to reproduce and keep a record of how it was developed. Also, because there were no statistical significance considerations in its construction, one cannot be certain this visually fit line does not introduce an error making the EW too large or too small. The method described here has the advantage of being based on a least squares fit to what is considered to be random noise.

The steps for determining the continuum and computing EW's and confidence limits are:

1. Identify the parts of the spectrum that are emission or absorption lines.
2. Remove the emission and absorption lines, i.e., setting their intensity values to system missing.
3. Develop a polynomial model to estimate the continuum from the remaining intensity values. These are considered to be noise.
4. Compute EW's and confidence limits.

The December 7, 2010 Sodium D lines spectrum of Epsilon Aurigae will be used as the demonstration spectrum. It was taken with a Meade 16" LX200R telescope, an LHIRES III spectrometer, and an ST-8XME camera at my observatory in Dewey, Arizona. Figure 1 is an SPSS plot of the profile from a .txt file written out by VSPEC. As mentioned, the usual protocol is to normalize a spectrum, i.e., dividing each intensity value by a continuum value. Figure 1 is not normalized, to demonstrate that normalization is not required to calculate an EW.

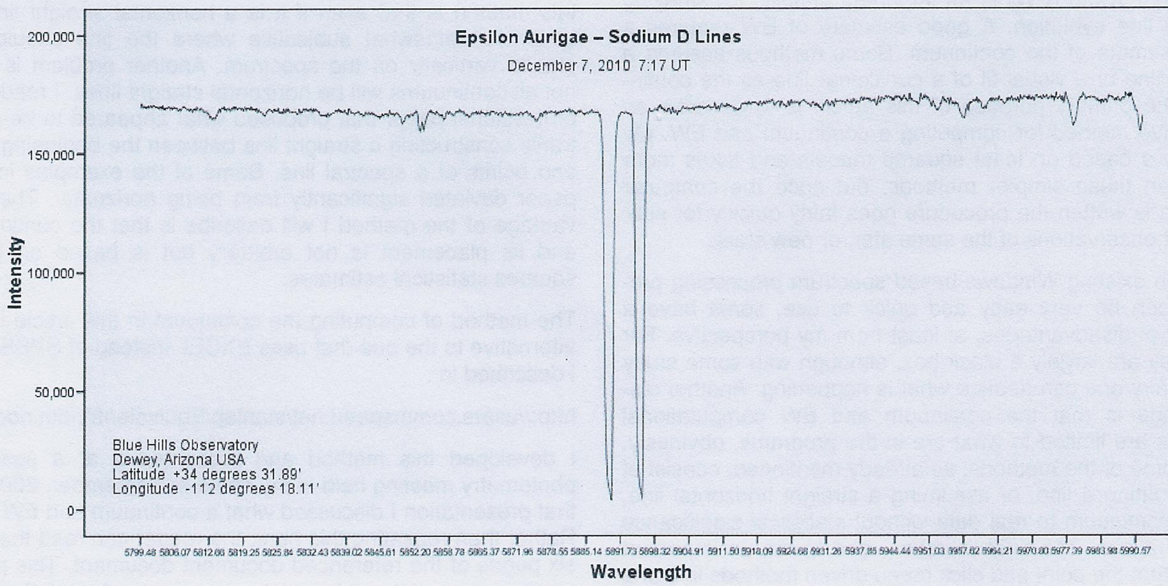

Figure 1

For this article I am fitting a continuum to the entire spectrum even though I am only interested in the Sodium D lines. From my experience, it is not always possible to use the entire spectrum and get a good continuum for a specific line. Nor do I think it is good, from a modeling perspective, to use the entire spectrum if different regions of the spectrum are not comparable. For example, if one is developing a model of Arizona's (Arizona is where I live) economy to predict some measure of economic growth, it would not make sense to include data going back to the start of the 20th century because the economy and demographics were entirely different then. One would probably have to add variables not relevant to today's economy. Likewise with a spectrum. It can be seen in Figure 1 that the spectrum is not entirely homogeneous. Instead of using the entire spectrum, as I am doing for this illustrative example, it may be more appropriate to use the left 2/3 for the Sodium D lines continuum estimate (although for this spectrum and lines it has virtually no impact). The right 1/3 has a cluster of lines and some dips the left 2/3 doesn't have. Consequently, it may be more appropriate to limit a continuum model for, say, the 5957A and 5978A lines to this region. Also, if the series is heteroscedastic (this one is not), by using the entire spectrum one would inflate or deflate the estimates of the upper and lower confidence limits of a line's EW depending on where the line is in the spectrum. If the line was in a region of low error, the confidence limits would be inflated. If it was in a region of high error, they would be deflated. Heteroscedasticity is another reason for dividing the spectrum into regions.

VSPEC is convenient to use to identify the beginning and end of absorption and emission lines. The beginning and end of the lines were entered into an SPSS computer program to remove them. Figure 2 is an SPSS plot of the spectrum with the lines removed. The remaining is considered to be noise.

This points to one potential pitfall of this method. Determining the beginning and ending of a line, what is a line, and the adjacent parts of the spectrum that may be affected by a line are subjective to a degree. The result may be that confidence limit estimates may be too small. On the other hand, relying on an automated method may result in estimates being too large.

The spectrum in Figure 1 is composed of 1461 pixels. Of these 710 were set to system missing values for intensity because the intensities were deemed to be absorption or emission lines, or influenced by lines. This left 751 intensity values as data points to develop a polynomial model of the continuum. A stepwise regression performed on the data resulted in the following model:

$$Y = A_1 + A_2X^2 + A_3X^3 + A_4X^4$$

where

Y = Intensity

X = Pixel location

A_1 = Constant term

A_2 to A_4 = Coefficients

It is easy to over fit a continuum model, i.e., keep adding higher order terms, X to the n^{th} power that, although statistically significant, only minimally reduce the residual error. In SPSS one can quickly fit regression models with higher order terms and then plot the standard error of a model against the order of the model. For the continuum being estimated, the plot in Figure 3 shows that terms of higher order than the fourth do not substantially reduce the residual error of the model. Hence, a fourth order model was accepted.

Figure 4 is a graph of the fitted line to the data. The sinusoidal shape of the curve is exaggerated by the scale of the graph. Figure 5 is a graph of the fitted line which is now called the continuum. As can be seen, the model very nicely fits the data and very well predicts the continuum values in the regions where the intensity values of the absorption and emission lines were set to system missing.

Figure 6 is the graph of the Sodium D1 and D2 lines with the continuum overlaid. For this spectrum the estimated continuum is essentially a straight horizontal line. Although there is a slight detectable slant to the continuum (it can be seen better in an enlarged graph), for all practical purposes this small deviation from a horizontal line would probably be insignificant for estimating an EW. One would than ask, Why go through all this trouble and not just assume a straight horizontal line? Three reasons. First, the continuum may not always be so close to being a horizontal line. Page 14 of

http://users.commspeed.net/stanlep/EquivalentWidth.doc

is one example where it is obviously not a straight line for H-α. Second, one still has the problem of where to place the continuum along the y axis. The least squares fit statistically places the line where it should be, and this can be better defended than an eyeball placement using point, click, sliding bar, and drag. Finally, one has a record of what was done that is reproducible.

The next step in the process is to determine the beginning and ending pixel locations for the two Sodium D lines from Figure 5. The D1 line starts at 671 and ends at 698. For the D2 line the respective values are 717 and 742.

The algorithm for computing EW is on pages 4 to 6 of the EquivalentWidth.doc document located at

http://users.commspeed.net/stanlep/EquivalentWidth.doc

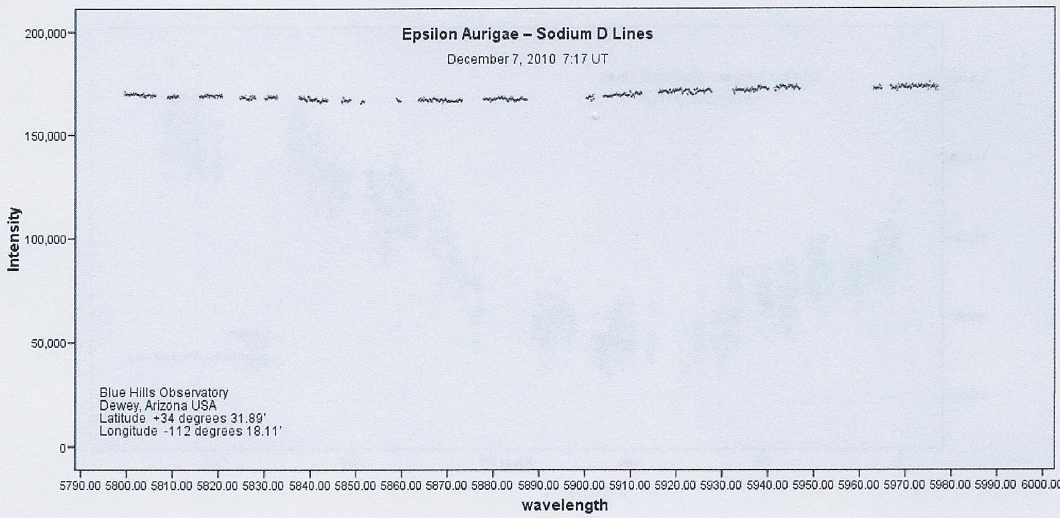

Figure 2

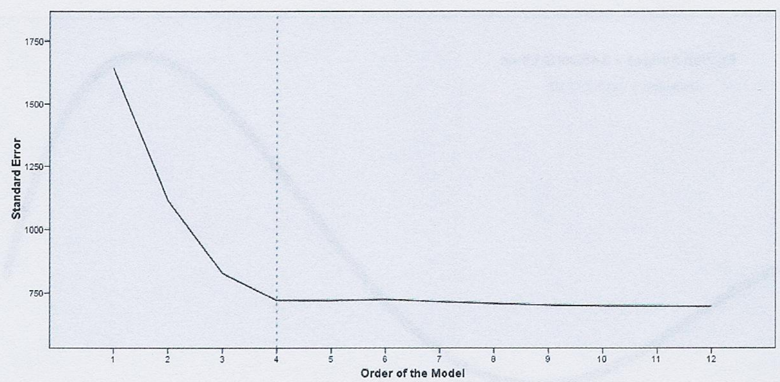

Figure 3

Essentially, it is the sum of the difference between the continuum, Ci, and intensity, Ii, divided by the continuum, and then multiplied by the width of a pixel in angstrom units, Δλ. The equation is:

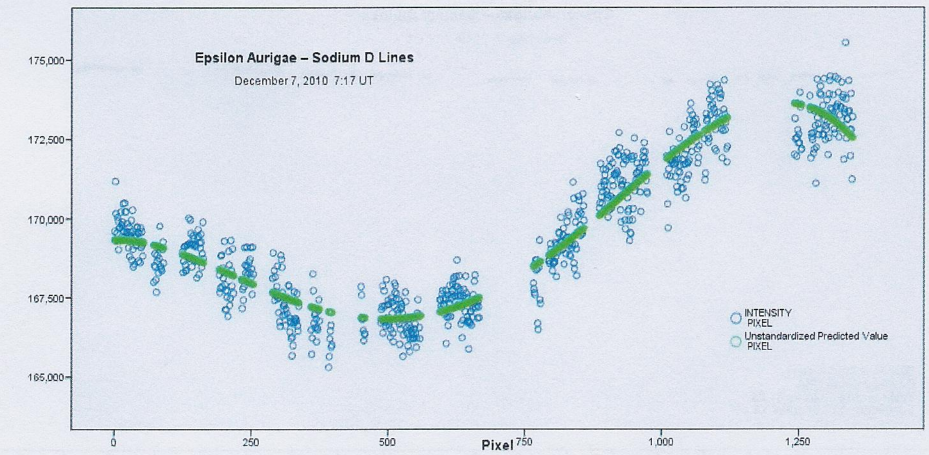

Figure 4

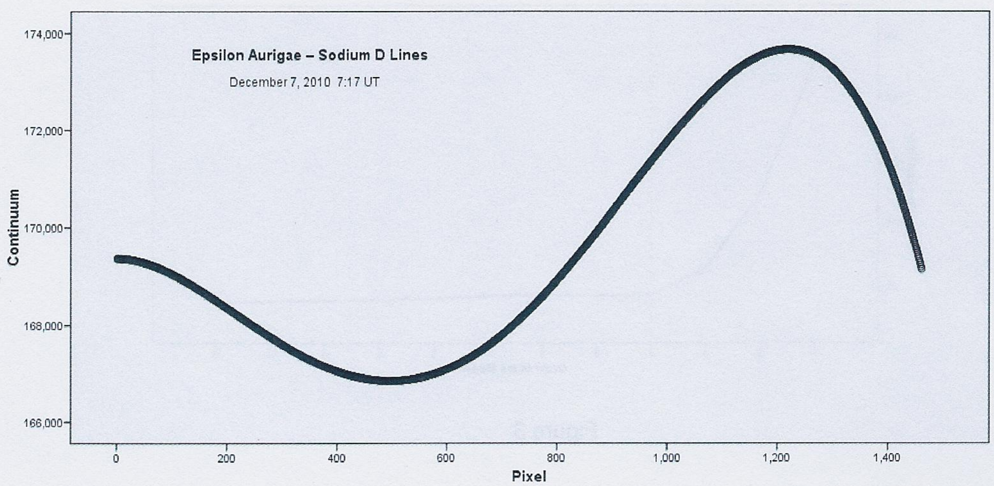

Figure 5

$$EW = \sum \frac{(C_i - I_i)}{C_i} \Delta\lambda$$

Δλ equals 0.132 A/pixel and was obtained by subtracting the angstrom value at the previous pixel location from the value at the current pixel location. Programming the EW equation and this lagged procedure for computing Δλ in an SPSS program, and entering the beginning and ending pixel locations for the two Na D lines, gives EW estimates for the D1 and D2 lines of 1.77Å and 1.66 Å, respectively.

The only thing left now is computing the upper and lower confidence intervals. How to do this is not described in books I have. I searched the internet but without success. I could have asked a professional or someone more knowledgeable but this would take the fun out of it. Read instructions? Heck no! After thinking about this for awhile it came to me. As already mentioned, the EW of a line is just the individual the contribution of each pixel, summed over all pixels. For one pixel, the contribution to the EW is

$$EW_{pixel} = \frac{(C_i - I_i)}{C_i} \Delta\lambda$$

where C_i and $\Delta\lambda$ have already been defined, the latter being equal to 0.132Å.

where σ is the standard deviation equal to 720.77, and t =1.96 for 750 degrees of freedom (recall there are 1461 pixels, 711 intensity values were set to system missing, and one degree of freedom is lost from the mean of the residu-

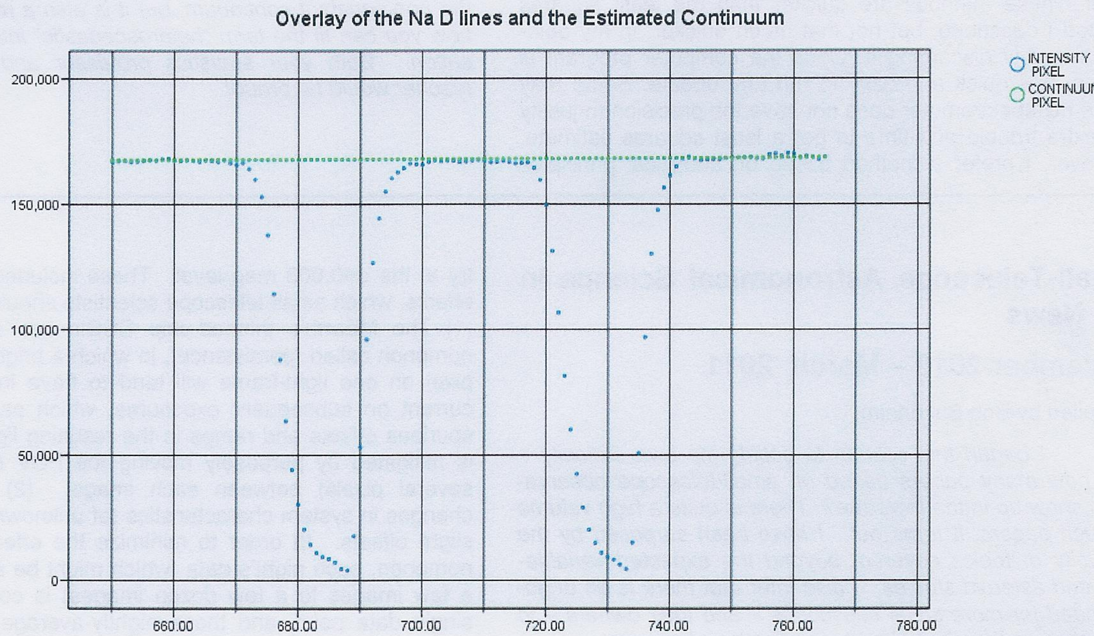

Figure 6

Each I_i is a random variable, i.e., it is the intensity of the spectrum at the i[th] pixel location, that deviates around the estimated continuum at the i[th] pixel location. Hence, to get a confidence limit estimate what is needed is an estimate of the standard deviation of the residuals of the regression model around the continuum. It is available in the SPSS regression output as the standard deviation of the residuals . In our example, itis equal to 720.77. It can be computed independently from the residuals, which is what I did.

From a corollary of a theorem in statistics, the variance of the sum of n independent variables is the sum of the variances of the n variables. In this case, the n variables are the n pixel values of the line intensities. If each I_i is treated as an independent variable, then the upper and lower confidence limits for each I_i are $(I_i + \sigma t)$ and $(I_i - \sigma t)$, respectively. Hence, computationally, the upper 95% confidence limit on EW s

$$UCL_{EW} = \sum \frac{[C_i - (I_i + \sigma t)]}{C_i} \Delta\lambda$$

The lower 95% confidence limit is

$$LCL_{EW} = \sum \frac{[C_i - (I_i - \sigma t)]}{C_i} \Delta\lambda$$

These algorithms were programmed into the SPSS program to get the following 95% confidence limit estimates.

	EW	Upper 95% C.L.	Lower 95% C.L.
D1 Line	1.77	1.80	1.74
D2 Line	1.66	1.69	1.63

Conclusion

This article demonstrated that some of the concepts and processing in spectroscopy are simple and not mysterious, and can easily be performed without having to rely on existing point and click software. It also provided an alternative method for determining the continuum, which is a crucial part of computing a good estimate of the EW. I have shown this can be done in statistical packages like SPSS, even EXCEL, and this allows one to keep a record of what was done so results can be reproduced. This is in contrast to some of the point and click and sliding bar methods of other, and admittedly easier to use, software. Finally, this article demonstrated that one can conduct, i.e., supplement, processing and analyses that might not be available in existing spectroscopic processing software.

I hope that this article has also demonstrated to the reader that there is an advantage to the least squares method over the visual fit methods. For example, one application of the latter is to visually fit a smoothed curvilinear line to a spectrum and then divide the spectrum by the smoothed line. This produces a spectrum that follows a nice horizontal line, straight as an arrow. However, there were no statistical significance considerations involved. Hence, as previously mentioned, one cannot be certain that this visually-fit line does not introduce an error, making the EW too large or too small. These methods are quicker than the least squares method I described, but not that much quicker, in my opinion, to justify discarding it. Once the computer program is written, it is quick and easy to run and update. Some may argue my spectrometer does not have the precision to justify the extra trouble and time to get a least squares estimate. However, I prefer a method based on statistical principles instead of guessing a line. I might abandon this for another method as I learn more about spectroscopy, but for now it works well for me.

For those interested, the SPSS syntax will be provided on request.

Editor's Note: Not only is this a powerful way to deal with the non-constant continuum, but it is also a rare example of how you can fit the term "heteroscedastic" into your conversation. Both your statistics professor and your English teacher would be proud!

Small-Telescope Astronomical Science in the News

December 2010 – March 2011

compiled by Bob Buchheim

I began this column to satisfy my own curiosity – just how many papers based on small-telescope observations show up in the literature? There is quite a high volume of such papers, it turns out. I have been surprised by the diversity of topics covered, beyond the expected variable-star and asteroid studies. I also infer that there is an ongoing need for more small telescopes – and their owners – to spend some time and effort in the service of science.

Since the response to this column has been positive, I'll continue it until either (a) a motion to cease is sent to the Newsletter editor, or (b) other obligations divert my attention. If you have comments on any of the research projects noted below, or other suggestion, please do let me know (rbuchheim@earthlink.net).

The GJ1214 Super-Earth System: Stellar Variability, New Transits, and a Search for Additional Planets
by Zachory K. Berta, et al
http://arxiv.org/PS_cache/arxiv/pdf/1012/1012.0518v1.pdf

This paper covers several different sets of data relating to the host star GJ1214 and its transiting exo-planet. The small-telescope portion comprises a long photometric study made with the MEarth telescope. MEarth is a set of 8 telescopes. Each telescope is 40 cm (16 inch) aperture, fitted with a CCD imager with 0.757 arc-sec pixels, and a 715 nm long-pass spectral filter (i.e. deep red and infrared). It turns out that it is of great value to accurately characterize the host star's photometry, so that large-telescope spectroscopic and planet-transit lightcurves can be properly interpreted. In particular, if the star's brightness varies, that variability can be transformed into random "noise" in spectra and transit-timing calculations. In order to characterize this star, the authors conducted a long-term (3 years) monitoring of the star's out-of-eclipse brightness. They found compelling evidence of a pseudo-periodic 52.7-day modulation, of amplitude ± 0.0035 mag (comparable to the transit depth) which is identified as the rotation period of the star (the modulation being caused by star-spots).

The authors also discuss several tricky features that they had to account for in order to be confident in their photometry at the $\sigma \approx 0.003$ mag level. These included the following effects, which small-telescope scientists should be aware of. (1) The MEarth's thinned-chip CCD's can exhibit a phenomenon called "persistence", in which a brightly-illuminated pixel on one light-frame will tend to have increased dark-current on subsequent exposures, which can give rise to spurious offsets and ramps in the resulting lightcurve. This is mitigated by purposely moving the FOV (by more than several pixels) between each image. (2) Night-to-night changes in system characteristics (of unknown origin) cause slight offsets. In order to minimize the effect of this phenomenon, each night's data (which might be anywhere from a few images to a few dozen images) is combined into a single data point, and these nightly-averages are used to search for the rotation-period signal. In addition, the array of nightly-averages is used to estimate the night-to-night photometric "jitter", and this "jitter" is treated as a random noise variable. (3) The target star is a very red star (an M-dwarf), whereas most available comp stars have more "normal" colors. This color difference between target and comp is modulated by night-to-night variations in atmospheric absorption in the water-bands that dominate the absorption in the spectral region being used; hence causing a sort of "second-order extinction" effect that can be comparable to the stellar-rotation lightcurve amplitude. In order to mitigate this effect, the team takes advantage of the MEarth observing sequence, in which the telescope cycles among several stars sequentially through the night. The ensemble of these observations is used to establish a nightly zero-point.

The key conclusions from this portion of the study are that GJ1214 is a spotted star with a now-known rotation period of 52.7 d, implying weak magnetic activity; that the level of star-spot photometric variability will not impede currently-feasible spectroscopic studies intended to learn about the planet's atmosphere (including HST and anticipated Spitzer observations); but that as higher-sensitivity spectrographs become available (e.g. the James Webb telescope), it will be necessary to simultaneously monitor the star's photometry to mitigate spurious effects in the spectra.

Qatar-1b: A Hot Jupiter Orbiting a Metal-Rich K Dwarf Star
by K. A. Alsubai, et al
http://arxiv.org/PS_cache/arxiv/pdf/1012/1012.3027v1.pdf

Small-telescope systems (such as HATnet, WASP, XO) have an enviable record of discovering exo-planets, but they