Differential Ensemble Photometry by Linear Regression

Kevin B. Paxson
20219 Eden Pines, Spring, TX 77379; kbpaxson@aol.com

Received January 20, 2010; revised January 22, 2010; accepted January 25, 2010

Abstract A linear regression method for multiple star ensemble photometry by spreadsheet is presented. After initial spreadsheet setup and data entry, a differential ensemble magnitude estimate is calculated along with a total error. Ensemble photometry by linear regression allows one to see the distribution of comparison star errors, yields variable star estimates with enhanced confidence, and identifies potential problems with the comparison stars, validating the Johnson *V* magnitude sequence. Spreadsheet construction is described and two linear regression ensemble examples are illustrated and discussed.

1. Introduction

Over the past decade and a half, CCD variable star photometry by amateur astronomers has grown exponentially with the increased availability of larger CCD chips and less expensive commercially-made CCD cameras in the marketplace. Photometric reduction techniques have become more refined over time, including ensembles. Most ensemble reduction methods involve the use of multiple sets of variable star-comparison star differences, where differential magnitude and error calculations are computed using a variety of methods. These include the mean value method, the weighted average method, and the master star method (Crawford 2006). Of these, the mean value method is most popular.

Ensemble photometry methods using mainframe computer programs (Honeycutt 1992) or specialized programming languages such as C (Richmond 2006) and FORTRAN (Howell and Everett 2001) exist in the literature, but are beyond the access of and availability to most amateur astronomers. Most image processing software packages (MAXIM DL, AIP4WIN V2, and others) can perform photometry of a single star, an ensemble, or photometry over an entire image. These software programs use the mean value method for ensemble determination, but do not allow for the visualization of an individual photometric result. The author used the single star photometry function in both MAXIM DL5 and AIP4WIN in V2 for the ensemble examples presented in this article.

Most spreadsheet applications have been oriented towards photometric corrections (Warner 2006), system transformations (Warner 2006), time series light curve analysis of variable stars (Cook 1999), asteroids (Warner 2006), or exoplanets (Gary 2007). The spreadsheet-based linear regression methodology for multi-star ensemble photometry was described in concept by Buchheim (2007) and has been expanded using Microsoft Excel 2007® in this paper.

2. The linear regression method of ensemble photometry

The linear regression method for ensemble photometry is an algebraic solution and graphical presentation in which instrumental magnitude and Johnson *V* magnitude pairs for all measure comparison stars are linearly regressed and plotted. The result is a straight line and a linear regression equation that relates Johnson *V* magnitude to instrumental magnitude with minimized residual error (see Figure 1). This equation takes the form of

$$Y_N = a + B \times X_N \quad (1)$$

where (Y_N) is the calculated Johnson *V* magnitude of comparison star N, (a) is a constant offset, (B) is the slope of the linear regression solution, and (X_N) is instrumental magnitude of comparison N. Excel or other spreadsheet programs can readily derive and graph the linear regression equation of the trend line for all Johnson *V* and instrumental magnitude pairs.

The coefficient of determination, or R^2, represents the statistical measure of how well the linear regression line represents the real data points. An R^2 value equal to 1 indicates that the regression line perfectly fits the observed data. With comparison star variation about the mean of the linear regression, R^2 is usually slightly less than 1, with the average distance from the line inversely related to R^2. R^2 usually increases with increased number of comparison stars and increasing image SNR.

The estimated magnitude error from the linear regression method is the mean of the differences between the calculated values from the linear regression and true Johnson *V* magnitudes for all of the measured comparison stars. Poisson photon noise error, which is caused by statistical photon distribution over a given time interval, is another source of error. Poisson photon noise error (for one sigma error) is approximated by Equation 2 (Howell 2000; Berry and Burnell 2005):

$$\text{Poisson noise error} = 1.0857 / \text{SNR} \quad (2)$$

The linear regression error is combined in quadrature with the Poisson photon noise error to arrive at a final standard error estimate. This final total error is given by Equation 3:

$$\text{Final total error} = \text{Square root}\,(\text{Linear regression error}^2 + \text{Poisson noise error}^2) \quad (3)$$

Specifics about the linear regression ensemble method and error calculation are explained in the V723 Cas spreadsheet example given below.

3. Spreadsheet ensemble example for V723 Cas

An eight-star photometry ensemble spreadsheet (Figure 3) was created for V723 Cas from a single 180-second Johnson *V* filtered exposure taken with a remote internet based 24-inch (0.61-m) f /10 Cassegrain. This spreadsheet was constructed with Microsoft Office Excel 2007® by using the following steps. When entering text in the steps below, enter the text between the quotation marks only, not including the quotation marks. Quotation marks may also indicate an action, a keystroke, or a computed result. In specific cells, all text is center-justified and all numbers and calculated results are right justified.

1. Please view Figure 2 and replicate this skeleton spreadsheet.

2. Format the empty cell J15 by right clicking on this cell, selecting "Format cell" and selecting "Perimeter highlighting" and selecting the lightest weight line and then hitting <ENTER>.

3. Select cell J15, right mouse click to "Copy" it and then "Paste" it in all cells which are surrounded by a black perimeter as shown in Figure 2.

4. Type in the specific text from each cell in Figure 2 into your spreadsheet.

5. In cells C6 to C13, type "132, 142, 145, 147, 150, 154, 159, and 162," respectively, hitting <ENTER> after each entry. These are the AAVSO chart comparison star *V* magnitudes to one decimal place for V723 Cas.

6. In cells D6 to D13, type "4.505, 5.550, 5.802, 6.045, 6.358, 6.732, 7.222 and 7.522," hitting <ENTER> after each entry. These are the instrumental magnitudes from your calibrated CCD image and your photometry software program for each comparison star.

7. In cells E6 to E13, type "13.203, 14.226, 14.465, 14.726, 15.033, 15.390, 15.877 and 16.165," hitting <ENTER> after each entry. These are the true Johnson *V* sequence magnitudes for each comparison star.

8. Highlight cells D6 through E13 (all 16 cells) and then go to the "Insert" column drop down menu and select a "Scatter Chart with Smooth Lines and Markers."

9. In the newly created graph below the data cells, right mouse click on the line between the data points and select "Add a Trendline" under Chart.

10. Select the "Linear" regression type and select "Display Equation on chart" and "Display R-squared on chart." Do not select "Set Intercept to 0.0" (all photometry software packages have their own specific zero magnitude offset). The resulting regression line equation in this newly created graph should read "y = 0.9838x = 8.7695" and "R^2 = 0.9999."

11. In cell F6, type "=0.983*D6+8.769" and hit <ENTER>. The result of "13.1974" is the calculated Johnson *V* magnitude for 132 comparison star.

12. Highlight cell F6 and then click the right mouse button to "Copy" this cell.

13. Highlight cells F7 to F13 and then right mouse click to "Paste" the regression equation in all these cells. The returned values in these cells are the calculated Johnson *V* magnitudes the remainder of the comparison stars.

14. In cell G6, type "=ABS(E6–F6)" and then hit <ENTER>. The result of "0.0056" is the error between the calculated and true Johnson *V* magnitude for the 132 comparison star.

15. Highlight cell G6 and then click the right mouse button to "Copy" this cell.

16. Highlight cells G7 to G13 and then right mouse button to "Paste" this equation into all these destination cells. The resulting values are the errors between the calculated and true Johnson *V* magnitudes for the remainder of the comparison stars.

17. In cell G14, type "=AVERAGE(G6:G13)" and hit <ENTER>. The result of "0.0072" is the average error between the calculated magnitudes and the true Johnson *V* magnitudes for each of the comparison stars.

18. In cell I6, type "6.428" and then hit <ENTER>. This is the instrumental magnitude of the V723 Cas from the calibrated CCD image and photometry reduction software.

19. In cell J6, type "=0.983*I6+8.769" and hit <ENTER>. The result of "15.0877" is the calculated linear regression value for the differential *V* magnitude of V723 Cas.

20. In cell I9, type "199.66" and then hit <ENTER>. This is the SNR reading of V723 Cas from the calibrated CCD image and the photometry reduction software.

21. In cell J9, type "=1.0857/I9" and then hit <ENTER>. The result of "0.0054" is the Poisson photon noise error (one sigma error in magnitude) of the V723 Cas SNR reading from the photometry reduction software.

22. In cell I12, type "=SQRT((J9^2+G14^2))" and then hit <ENTER>. The result of "0.0090" is the "Final *V* error." The final V723 Cas magnitude estimate is 15.0877 *V* (cell J6) and the error is 0.0090 (cell I12).

23. Type in photometry software and annulus or radii information (I typed "MaxIm DL5 and 6, 16") in cells I15 and J15.

24. Type in your specific details (I typed "V723 Cas, SSO 24", V, 180 and 2454801.6952") in cells F3, G3, H3, I3, and J3, hitting <ENTER> after each cell entry.

25. Save your file with a meaningful file name (for this example I called the Excel file "V723_Cas_01Dec2008.xls") and put it in your Ensembles folder.

26. This ensemble spreadsheet can be copied and renamed with the "File," and "Save As" options and later amended for any star in your observing program.

Please note that instrumental magnitudes by definition are negative, but most software packages express raw instrumental magnitudes as a positive number. These positive instrumental magnitudes are used in the spreadsheet.

4. The V723 Cas ensemble—results and methodology

The caption below Figure 3 explains the workings of the linear regression ensemble spreadsheet. Instrumental magnitude (X axis) and Johnson *V* (Y axis) pairs for all eight comparison stars are regressed and plotted, resulting in a very straight line with an X coefficient of 1.0373 and a very high R^2 (correlation coefficient) of 0.9999. The Johnson *V* magnitude estimate of 15.0877 with a total error of 0.0090 is an excellent photometric result. This is due to the high SNR (199.66) attained in the 180-second exposure taken by the 24-inch (0.61-m) f/10 Cassegrain.

5. Methodology

I typically use four to seven or eight comparison stars for my ensembles, ranging no more than two and a half magnitudes above and below the targeted variable's estimated magnitude. Due to crowded star fields and/or the necessity

of picking a constant measuring aperture and outer (sky) annulus for all stars to be measured on a CCD image, one may not be able to use all of the comparison stars in a given sequence. Use as many comparison stars as possible. Using three comparison stars or less is not recommended. Three data points are an insufficient sample for this ensemble methodology and any two points define a straight line.

A good linear regression ensemble should yield a regression equation with an X coefficient of 0.985 to 1.015 and an R^2 value of 0.99 or greater. R^2 usually increases and comparison star data scatter usually decreases and with an increasing number of comparison stars and image SNR. Look for any outliers (0.1 magnitude or greater) that plot away from a straight line beyond average scatter. If measured SNR is high and no signs of comparison star contamination can be found, you may have discovered a potential variable star! If the brighter stars in a given sequence make the ensemble graph in a "hockey stick down" manner, pixel saturation may be the culprit.

Should a calculated Johnson *V* magnitude for one comparison star substantially fall off the linear regression line, that star should be considered suspect. Saturated pixels, hot or cold pixels, dust accumulation since the last flat was taken, cosmic ray hits in a comparison star image, or potential comparison star variability may be possible explanations for comparison star error. If all of the other comparison stars fall on a straight line, the one bad comparison star can be deleted with confidence. This procedure is explained below.

In the spreadsheet for V723 Cas shown in Figure 3, the comparison stars 4 and 5 (147 and 150) have the highest error. For illustrative purposes only, we will eliminate the comparison star 4 (147) and recalculate the ensemble result. Begin by highlighting cells B9 to G9 and hit <DELETE>. Highlight cells B10 to G14 and move them up one row vertically. A new equation is now in the regression graph, now reading "y = 0.9841 * x + 8.7663" and the "R^2 = 1" versus the old equation which read "y = 0.9838 * x = 8.7695" and the "R^2 = 0.9999." In cell F6 type "=0.9841*I6+8.7663" and hit <ENTER>. "Copy" cell F6 and highlight cells F7 to F12 and right mouse click to "Paste" this new equation into these destination cells. All of the cells from F6 to F12 have new Johnson *V* magnitudes based on this new linear equation. Linear regression errors in cells G6 to G12 have also been updated. In cell J6, again type the equation "=0.9841*I6+8.7663" and hit <ENTER>. The recalculated ensemble magnitude for V723 Cas using seven comparison stars is now 15.0921 *V* (versus 15.0877 *V*) and the total error is now 0.0073 (versus 0.0090).

6. The 3C 66A ensemble—results

A nine-star photometric ensemble was created for the blazar 3C 66A from a single 180-second clear filtered exposure taken with a remote internet-based f/3.8 10-inch (0.25-m) astrograph. For brevity, the spreadsheet ensemble with all

of the photometric and sequence data values already entered into the appropriate cells is shown in Figure 4.

This linear regression ensemble is of good quality with an average X coefficient of 1.0373 and an R^2 value of 0.9994, especially when you consider this is an unfiltered image. The Johnson *CV* magnitude estimate of 14.820 with a total error of 0.031 is respectable. Comparison stars 8 (15.872 *V*) and 9 (16.520 *V*) have the highest error, due to decreasing SNR (17.808 and 10.988, respectively). These two comparison stars could be ignored for a better ensemble result if desired.

7. Conclusions

A linear regression based-method for ensemble differential photometry has been presented and described. By visualizing the distribution of the errors of the comparison stars in a sequence, confidence is increased in the variable star estimate. The linear regression methodology provides an objective method for ensemble magnitude estimates and to handle total error correctly. Good linear regression ensembles validate the accuracy of a variable star sequence, more so than simpler ensemble methods. A comparison star "outlier" that dramatically departs from the linear regression can be readily identified and investigated. If image quality sources of error are eliminated, the outlier may be a variable star. The linear regression ensemble method forces the observer to "look at the data," and access the quality of the sequence, resulting in a confident variable star magnitude estimate.

8. Acknowledgements

The author would like to recognize Arne Henden (2009) for his email suggestions concerning this methodology offered in late 2008 and Mike Simonsen and others for their comments during the preparation of this paper.

References

Berry, R., and Burnell, J. 2005, *Handbook of Astronomical Image Processing*, Willmann-Bell, Richmond, VA, 278.

Buchheim, R. K. 2007, *The Sky Is Your Laboratory*, Springer Praxis, Berlin, 107–112.

Cook, L. L. 1999, AutoPlot—A Light Curve Plotting Excel Spreadsheet, http://www.lewcook.com/autoplot.htm.

Crawford, T. R. 2006, Ensemble Photometry Options for AAVSO Amateur Observers Using AIP4WIN and Other Software, http://homepage.mac.com/windwalker1/.Public/Ensemble%20PhotometryV5.pdf.

Gary, B. L. 2007, Exoplanet Observing for Amateurs, http://brucegary.net/book_EOA/WebEOA/ExoplanetObservingAmateurs.html.

Henden, A. A. 2009, Private communication.

Honeycutt, R. K. 1992, *Publ. Astron. Soc. Pacific*, **104**, 435.
Howell, S. B. 2000, *Handbook of CCD Astronomy*, Cambridge Univ. Press, Cambridge, 57.
Howell, S. B., and Everett, M. E. 2001, *Publ. Astron. Soc. Pacific*, **113**, 1428.
Richmond, M. 2006, Ensemble Software to Perform Inhomogeneous Ensemble Photometry, http://spiff.rit.edu/ensemble.
Warner, B. D. 2006, *A Practical Guide to Lightcurve Photometry and Analysis*, Springer, New York.

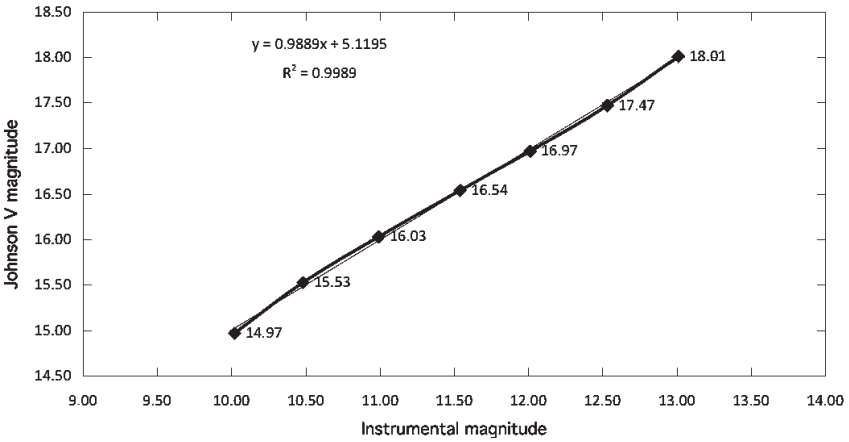

Figure 1. A sample graph showing the linear regression concept. X and Y pairs of Instrumental (X-axis) and Johnson *V* (Y-axis) magnitudes of all the comparison stars are put into a linear regression, yielding an equation which relates the pairs. In this particular example, Johnson *V* magnitude = 0.9889 × Instrumental Magnitude + 5.1195. The R^2 is 0.9985.

	A	B	C	D	E	F	G	H	I	J
1										
2						Variable	Telescope	Filter	Exp. In sec.	JD
3										
4										
5		Comp Star #	V m.	Instr. m.	V m.	Calc. V m.	Error V m.		Var. Inst. m.	Var. V m.
6		1								
7		2								
8		3							Var. SNR	Std. Error
9		4								
10		5								
11		6							Final V Err.	
12		7								
13		8								
14						Avg. V error			Software	Apertures
15										

Figure 2. The linear regression ensemble spreadsheet skeleton. Replicate this spreadsheet and follow the instructions in the text of Section 3 to recreate the spreadsheet and graph shown in Figure 3.

	A	B	C	D	E	F	G	H	I	J
1										
2						Variable	Telescope	Filter	Exp. In sec.	JD
3						V723 Cas	SSO 24"	V	180	2454801.6952
4										
5		Comp Star #	V m.	Instr. m.	V m.	Calc. V m.	Error V m.		Var. Inst. m.	Var. V m.
6		1	132	4.505	13.203	13.1974	0.0056		6.428	15.0877
7		2	142	5.550	14.226	14.2247	0.0014			
8		3	145	5.802	14.465	14.4724	0.0074		Var. SNR	Std. Error
9		4	147	6.045	14.726	14.7112	0.0148		199.66	0.0054
10		5	150	6.358	15.033	15.0189	0.0141			
11		6	154	6.732	15.390	15.3866	0.0034		Final V Err.	
12		7	159	7.222	15.877	15.8682	0.0088		0.0090	
13		8	162	7.522	16.165	16.1631	0.0019			
14						Avg. V error	0.0072		Software	R1 and R2
15									MaxIm DL5	6, 16

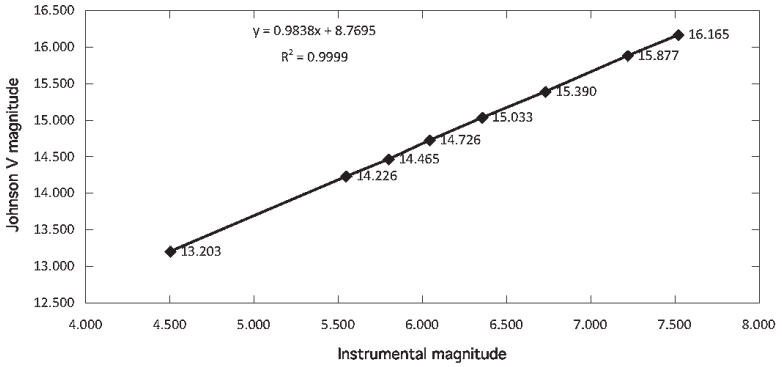

Figure 3. A linear regression ensemble spreadsheet for V723 Cas. Instrumental magnitudes (Column D data and plotted on the X-axis) and Johnson *V* magnitudes (Column E data and plotted on the Y-axis) are regressed for all of the measured comparison stars to yield a linear equation. The Johnson *V* magnitude estimate for V723 Cas = 0.9838 × 6.428 (the instrumental magnitude for V723 Cas given in cell I6) + 8.7595 = 15.0877 (cell J6). Final total error is ±0.0090 (cell I12). V723 Cas was not displayed on the graph in Figure 3 above, as only the comparison stars were regressed to establish the linear Instrumental to Johnson *V* magnitude relationship. V723 Cas would plot exactly on the linear regression line at X = 6.428 (instrumental magnitude) and Y = 15.0877 (the calculated Johnson *V* magnitude).

	Variable	Telescope	Filter	Exp. In sec.	JD
	3C 66A	GRAS-005	V	180	2454806.65492

Comp Star #	V m.	Inst. m.	V m.	Calc. V m.	Error V m.
1	129	17.529	12.856	12.831	0.025
2	136	18.311	13.639	13.642	0.003
4	142	18.861	14.182	14.212	0.030
5	144	19.074	14.424	14.433	0.009
6	148	19.392	14.771	14.763	0.008
7	153	19.915	15.308	15.306	0.002
8	159	20.505	15.872	15.918	0.046
9	165	21.042	16.520	16.475	0.045
				Avg. V error	0.021

Var. Inst. m.	Var. V m.
19.447	14.820

Var. SNR	Std. Error
46.112	0.235

Final V Error
0.236

Software	Apertures
AIP4WinV2	7, 10, 17

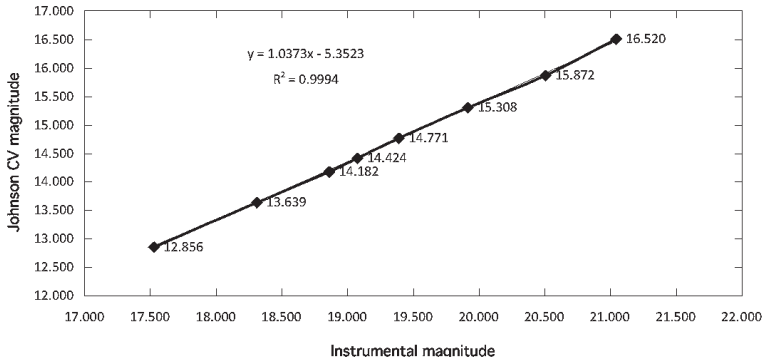

Figure 4. A linear regression ensemble spreadsheet for 3C 66A. The Johnson *CV* magnitude estimate for 3C 66A = 1.0373 × 19.447 (the instrumental magnitude for 3C 66A given in cell I6) −5.3523 = 14.820 (cell J6). Final total error is ±0.031 (cell I12). 3C 66A was not displayed on the graph in Figure 4 for reasons previously explained. 3C 66A would plot exactly on the linear regression line above at X = 19.447 (instrumental magnitude) and Y = 14.820 (the calculated Johnson *CV* magnitude).