Accuracy and Precision in Amateur Photometry

Edward O. Wiley
Live Oaks Observatory, Stellar Skies, Pontotoc, TX; ewiley@suddenlink.net

Kenneth Menzies
Tigh Speuran Observatory, Framingham, MA; kenmenstar@gmail.com

Received December 5, 2021; revised February 5, 2022; accepted February 7, 2022

Abstract Photometric accuracy and photometric precision were determined using the average magnitude and standard deviation of 7 to 10 images of 63 Landolt stars taken from 11 Northern Landolt fields by two amateur observers using CCD sensors in B, V, and I_c. Similar measures were taken for two of these Landolt fields using a CMOS sensor from the AAVSO Bright Star Monitoring NH2 observatory. A series of analyses were performed on observed average magnitudes compared to known Landolt magnitudes of the pooled data under different treatments that included both transformed and untransformed analyses under both single comparison star and ensemble treatments using observed minus known magnitude values (O–K analysis). A variety of non-parametric tests of magnitudes resulting from different treatments using absolute O–K values was used to assess the statistical differences between treatments. Regression analysis using untransformed ("raw") O–K values and B–V color indexes for each star were used to assess the differences between transformed and untransformed treatments for each filter and test for any statistical differences. Correlation analysis was used to assess the relationship between accuracy and precision. In most cases, transformed magnitudes are statistically more accurate than untransformed magnitudes. Even when there is no statistical difference in median values between transformed and untransformed results there is a statistically significant difference in the regression analysis indicating that transformation improves accuracy for the data as a whole in each filter. There were no statistical differences between the 16-bit CCD results and the 12-bit CMOS results for the two fields analyzed. Both were capable of a median accuracy of 0.02 magnitude or less, which is similar to the accuracies of the same APASS secondary standard stars in four of the fields included in the study. We detected no statistical difference between using a single versus small ensemble of comparison stars but prefer ensembles for reasons given. Precision is not correlated with accuracy nor need it be for some studies.

1. Introduction

Amateur photometrists desire their measurements to be both accurate and precise. But what does this mean? Photometric magnitudes reported to AAVSO are generally reported with uncertainty values, either derived from signal-to-noise estimates or standard deviation of the target star (if slowly varying), a check star of similar magnitude, or an ensemble. These estimates are certainly a measure of uncertainty of submitted observations, but how do they relate to the accuracy of the observations? To carefully assess uncertainty, we need to clearly separate accuracy (i.e., systematic error) and precision (i.e., random error).

Mandel (1964) outlines two concepts of accuracy: (1) accuracy relative to a value accepted as the "real" value, or (2) a value assigned to be true by consensus (or agreement). The value of the speed of light in a vacuum is an example of (1). A value assigned by expert consensus, as in the value of the meter, is an example of (2). Landolt standard stars (LSS) are an example of magnitude values assigned by expert opinion (2). Therefore, the difference between a measured magnitude of a Landolt star and the Landolt standard magnitude will provide a measure of accuracy within the accepted value of the uncertainty of the Landolt standard magnitude.

Precision is harder to define. To Mandel, it is easier to define imprecision: "Given a well described experimental procedure for measuring a particular characteristic of a chemical or physical system, and given such a system, we can define imprecision as the amount of scatter exhibited by the results obtained by repeated application of the process to that system" (Mandel 1964, p. 103). For example, if you repeatably measure the brightness of a slowly changing variable or a check star, then the standard deviation will provide a measure of imprecision. In photometry, imprecision is often referred to as a measurement of uncertainty. In general, it is meant to describe the distributional scatter of point source measures in a (hopefully) Gaussian set of observed magnitudes.

Papers in the literature with discussions of accuracy and precision fall into two categories. In the professional literature the concern is accurate measurement of flux by careful control of image acquisition and processing under known conditions (e. g., Stubbs and Tonry 2006). When Stubbs and Tonry (2006) use the term accuracy, they refer to the accuracy of uncertainty values of flux measurements. In a similar vein more applicable to amateurs are papers outlining best practices in photometry that are likely to improve precision (e.g., Newberry 1999; Koppelmann 2005). Sonnett *et al.* (2013, p. 446) define a measure of photometric accuracy as the Root Mean Square (RMS) residual of a magnitude estimate from a light curve model. An assessment of fit with other observations to a light curve model has merit in identifying outliers, but our understanding of models is that they are never true; their function is to predict future observations. Thus, they do not fulfill the accepted concepts outlined by Mandel. We propose to address accuracy and imprecision on the level of amateur photometry directly by addressing the Mandel criterion, comparing a result to a known standard.

Our main objective is to access photometric accuracy using differential aperture photometric techniques with typical amateur equipment and protocols. In doing so we hope to provide protocols for other amateurs to access their accuracy by imaging Landolt (or other) standard star fields. The ability to produce reasonably accurate results using standard stars gives confidence that measures of variable stars are also reasonably accurate in spite of the fact that no one can access the true accuracy of a variable at any given time of observation.

We assess accuracy using observed and known BVI_c magnitude values of Landolt Standard Stars (LSS) by comparing their known accepted magnitudes (K) against their observed magnitudes (O) using O–K analysis, a variant of O–C analysis using the known magnitude rather than a magnitude computed from a model. Magnitudes reported by Landolt (2013) were derived by repeated measures over several nights. The uncertainties reported are "mean errors of means," and not directly comparable to uncertainties of a single nightly measure or the standard deviation of a series of measures. Thus, to directly compare our accuracy with the Landolt standard would require observations over multiple nights, a research method not likely to be employed by amateur photometrists. However, Landolt (1983, p. 450) provides a method to calculate the "mean error for a single observation" by multiplying the square root of the number of nights the star was observed by the mean error of means. We performed these calculations on one field comprising all the stars of SA20 for B, V, and I_c magnitudes. Johnson V single observation error ranged from 0.001 to 0.011 magnitude: B from 0.001 to 0.011 magnitude, and I_c from 0.002 to 0.018 magnitude. These ranges are accuracy ranges, not precision ranges as the mean-of-means magnitude is taken as the known standard value. We conclude that any magnitude that we might measure, that is within 0.01 O–K, would be considered very accurate.

We evaluate various data reduction approaches for accuracy of magnitude estimates under transformed and untransformed protocols as well as for single comparison star versus ensemble comparison star protocols. Each such recalculation of the data is referred to as a "treatment." We ask four questions about both accuracy and precision:

1. Does transforming data into the standard Johnson-Cousins magnitude system (BVR_cI_c) using differential photometric protocols improve the accuracy of magnitude estimates?

2. If so, what is the effect on accuracy if we use more than one comparison star to form a small ensemble?

3. What is the relationship between accuracy and precision?

4. What differences are there between magnitude estimates made with two 16-bit CCD sensors and those taken with one 12-bit CMOS sensor?

2. Equipment

The following systems were utilized to conduct this study:

(a) Live Oaks Observatory (LOO). Location: 30.98° N 98.94° W. Mount: AstroPhysics Ap900 (German Equatorial). OTA: Celestron on HD with focal reducer, 280 mm f/7. Detector: Moravian G21600 Mk.1 (1536 × 1024 pixels, 9-micron square pixels, bin 1). Filters: B,V,I_c. Flats: Light box. Capture software: PD Capture. Reduction software: LesvePhotometry (de Ponthière 2011) Field of View: 24 × 16 arcminutes.

(b) Tigh Speuran Observatory (TSO). Location: 42.31° N 71.42° W. Mount: Paramount ME (German Equatorial). OTA: Hyperion, 317 mm f/8. Detector: SBIG STL-6303e (3072 × 2048 pixels, 9-micron square pixels, bin 2). Filters: B,V,I_c. Flats: Sky. Capture software: Maxim DL. Reduction software: LesvePhotometry. Field of View: 37 × 25 arcminutes.

(c) BSM_NH2 Observatory (BSM-NH2). Location: 43.69° N 71.56° W. Mount: Paramount ME (German Equatorial). OTA: Takahashi Epsilon, 180 mm f/2.8. Detector: ZWO ASI-183 (5496 × 3672 pixels, 2.4-micron square pixels, bin 2, gain 0). Filters: B,V,I_c. Flats: Sky. Capture software: Maxim DL. Reduction software: LesvePhotometry. Field of View: 90 × 60 arcminutes.

3. Methods

Five Landolt fields were imaged at the LOO observatory and six Landolt fields were imaged at the TSO observatory, both using CCD imagers. Two Landolt fields were imaged at the BSM NH2 Bright Star Monitor Network observatory using a CMOS imager. Details of each field observed are shown in Table 1. Ten images were taken of each target field and calibrated using dark, flat, and bias frames. Acceptable images were uploaded to LesvePhotometry for analysis, resulting in 7 to 10 images of each standard field. In LesvePhotometry standard field star magnitudes were downloaded from the AAVSO VSD comparison star database via the AAVSO VSP chart-creation software. For each field, a surrogate target star that was not a Landolt Standard Star (LSS) was selected as the target; the resulting sequences (LSS comparison stars and target) were saved as a master sequence in an Excel® workbook. The surrogate target was not analyzed but used as a place holder required by LesvePhotometry.

LesvePhotometry (LP) uses terms differently than AAVSO. The AAVSO "comparison" star is designated in LP with "R" (reference). An ensemble of these comparison stars would all be labeled "R" in LP but labeled as a comparison star ensemble in AAVSO nomenclature. The check star is the same in both nomenclatures. However, in LP we can introduce additional

Table 1. Observatories, detectors, and image fields.

Observatory	*Detector*	*Field*	*Date*
L00	CCD	SA20-SF4	1/18/2020
L00	CCD	SA32-SF1	11/24/2019
L00	CCD	SA26-SF1	2/25/2020
L00	CCD	SA95 (SW)	1/19/2020
L00	CCD	SA98-SF1	2/24/2020
TSO	CCD	SA20-SF2	2/20/2020
TSO	CCD	SA23-SF1	2/21/2020
TSO	CCD	SA23-SF4	4/24/2021
TSO	CCD	SA26-SF1	2/23/2020
TSO	CCD	SA32-SF4	4/24/2021
TSO	CCD	SA38	6/8/2020
BSM_NH2	CMOS	SA38	7/7/2020
BSM_NH2	CMOS	SA32-SF4	4/24/2021

stars that would function as additional check stars by designating these available comparison stars with the LP designation "C." In a Landolt field this allows for several standard stars to function as "targets" for magnitude estimation.

A number of templates were produced from the master sequence with different combinations of standard stars to be treated as Landolt targets for magnitude estimation (CK and C stars in LP terms) and one or more stars to be treated as comp stars (R in LP). Each combination is herein termed a "treatment." Transformed treatments were labeled with "-T," and untransformed treatments with "-NT." We use the more familiar AAVSO terminology, so "1C-NT" refers to a treatment with one comparison star (R), one check star (CK), and a variable number of additional Landolt standard stars as "target" check stars (C). Treatments were as follows:

(a) 1C: single comparison star (R), one check (CK), and all remaining LSS as additional "target" check stars (C). Both a transformed treatment (1C-T) and untransformed treatment (1C-NT) analysis were performed. They were applied to all image sets.

(b) 2C: two-star comparison ensemble (R), one check (CK), and all remaining LSS as additional "target" check stars (C). As above, both transformed and untransformed treatments were performed (e.g., treatments 2C-T, 2C-NT). Two iterations of 2C analysis were performed, switching two reference comparison stars for two comp stars (e.g., switching two "R" comps to two "C" comps and vice versa) to increase sample size. They were applied only to the LOO image sets.

(c) allC: All but two LSS as an ensemble (R), with one check and one additional "target" check star (C). The allC analyses were iterated so that each LSS, in turn, was a "target," one was a check, and the remaining were a comparison ensemble (Rs). So, if the field comprised five LSS, there were five analyses. This analysis was only performed on transformed data for reasons given in the discussion and applied only to the LOO image sets.

(d) 3C: Ensemble of three comparison(R), one check (CK), and all remaining LSS as "target" check stars (C). Both transformed and untransformed analyses (3C-T, 3C-NT) were performed, and applied to both TSO and BSM-NH2 image sets, including CCD and CMOS images, respectively.

Differential aperture photometry was performed in LesvePhotometry. Results were sorted in Excel® spreadsheets by BVI_c filter and treatment (T, NT). The magnitudes of target stars (CK and C stars) were averaged (N = 7–10, depending on image quality) and the standard deviations were calculated as a measure of precision. In addition to our observations, AAVSO Photometric All-Sky Survey (APASS; Henden *et al.* 2018) standard magnitudes are known for four of the Landolt Standard Fields (SA20-SF4, SA23-SF1, SA95, and SA98) for both the Johnson B and Johnson V bandpasses. We performed O–K analysis of the APASS magnitudes to compare to our own results.

Statistical tests, regression analysis, and boxplot visualizations of central tendencies and variation were conducted under the assumption that the measures of stars in the same and different fields for each filter could be combined into a single population of measures. Data were organized by filter and treatment. Statistically significant differences are denoted by an asterisk (*) in the tables.

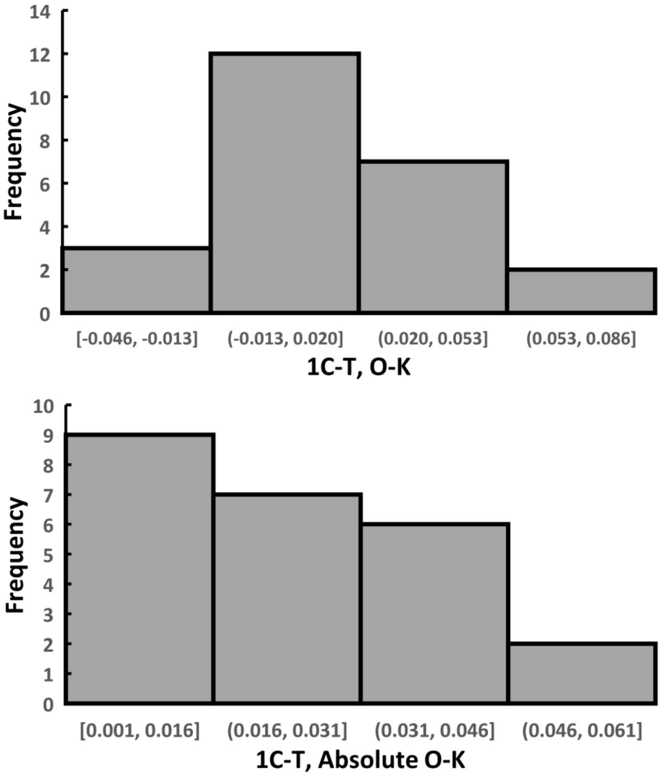

Figure 1. Frequency distributions of Johnson B filter results from Live Oak Observatory. Upper plot: frequency distribution of the O–K values of the One Comparison Star transformed analysis (1C-T). Lower plot: the same data except the O–K values have been transformed to absolute values.

The questions concerning accuracy of observed measures were addressed using observed minus known (O–K) magnitude analyses similar to the more familiar O–C (observed minus calculated) analyses. Values of O–K may be positive or negative. Untransformed ("raw") O–K values may lead to spurious estimates of accuracy since O–K = –0.1 and O–K = +0.1 average to a perfect agreement/accuracy of O–K = 0 when, in fact, both estimates are off by 0.1 magnitude. Thus, absolute values of O–K are more appropriate to assess accuracy. However, transforming O–K values to absolute O–K values sometimes resulted in data distributions that were not normally distributed (Figure 1a, b). Because of this we chose to express central tendencies as medians and adopted a non-parametric approach to evaluate the equality of median absolute O–K values for different treatments.

We used box plots prepared in Excel® to visualize medians and interquartile ranges of different treatments chosen for statistical treatment. Median and interquartile range (IQR) were conducted using a convenient on-line calculator (https://www.calculatorsoup.com).

Statistical comparisons consisted of analyzing the difference between different treatments under the null hypothesis that median values of absolute O–K were statistically identical. However, these data are frequently not normally distributed, and they are also highly correlated, that is, consisting of data of the same star under different treatments. Because of this, we evaluate the null hypothesis that the observed absolute O–K values between treatments are statistically identical using the Wilcoxon Signed Rank test, a non-parametric test designed to evaluate the effects of treatment when the samples are correlated, as is the case here. Where sample size was different between treatments (e.g., 1C data versus 2C data), restricted matrices were prepared that contain only stars common to both analyses. This meets a necessary condition of the Wilcoxon test that requires paired data. Otherwise, all stars were included for 1C-NT versus 1C-T. The null hypothesis is that the medians of the absolute O–K values and their distributions were statistically similar at the p = 0.05 significance level. The tests were conducted on a convenient on-line calculator (https://www.socscistatistics.com).

Questions about the effects of transformation are addressed with untransformed ("raw") O–K values and regression analysis in Excel®. The B–V color index was designated as the independent variable and the O–K value was designated as the dependent variable. In each treatment the significance of the B–V color index to predict the O–K value was taken as the effect of transformation under the hypothesis that a significant lack of prediction (acceptance of the null hypothesis that there was no association between B–V and O–K, p = >0.05) indicates the positive result of a successful transformation to the Standard Magnitude System. For example, if 1C-NT rejects the null and a 1C-T accepts the null, this indicates that transformation is effective. How effective is a matter of each individual measure, but the overall effect can be judged by the slope of the least-squares fit. A perfect transformation would result in a flat (zero slope) least-squares fit along the O–K = 0 axis. Lack of independence prevents further tests.

Precision was determined by a correlation analysis of absolute O–K and the standard deviation of the mean value of 7 to 10 individual measures of each target star. Correlation analysis was conducted using regression analysis in Excel® where the "Multiple R" value is the correlation coefficient, and a significant value is returned.

4. CCD accuracy and precision—results and discussion

Absolute O–K medians and variation around the median for each treatment are reported in Table 2 and visualized in Figure 2. The general trend is for untransformed data to be less accurate (i.e., larger O–K) and have greater variation (i.e., broader IQR) than transformed data. Most obvious findings were the Johnson B results where both LOO and TSO data show significant differences between untransformed absolute O–K averages (range 0.024–0.076) compared to transformed averages (0.012–0.019). A similar, albeit less dramatic, difference was noted for Cousins I_c. TSO Johnson V also showed improvement for the TSO analyses, but LOO Johnson V showed little to no transformation effect. To informally compare our results to APASS secondary standards we also show box plots of median, quartile, and range variation of transformed Johnson B and V magnitudes of APASS stars from four of the fields that had those data (Figure 2). We note that our transformed data compare well with the APASS data.

The average precision is also reported in Table 2. Precision estimates do not show any obvious differences between transformed and untransformed accuracy estimates. The range of standard deviation among both transformed (0.006–0.019)

Table 2. Summary data for combined standard fields, CCD.

Obs.	*Filter*	*Treatment*	*Median Abs. O–K*	*IQR O–K*	*Precision*	*N*
LOO	B	1C-NT	0.062	0.092	0.013	24
LOO	B	1C-T	0.019	0.031	0.016	24
LOO	B	2C-NT	0.076	0.081	0.013	24
LOO	B	2C-T	0.014	0.023	0.019	29
LOO	B	allC	0.021	0.023	0.012	29
LOO	V	1C-NT	0.011	0.01	0.008	24
LOO	V	1C-T	0.01	0.009	0.009	24
LOO	V	2C-NT	0.001	0.017	0.006	29
LOO	V	2C-T	0.009	0.016	0.006	29
LOO	V	allC	0.001	0.016	0.01	29
LOO	I_c	1C-NT	0.024	0.031	0.01	25
LOO	I_c	1C-T	0.005	0.014	0.011	25
LOO	I_c	2C-NT	0.011	0.085	0.008	28
LOO	I_c	2C-T	0.001	0.039	0.009	29
LOO	I_c	allC	0.002	0.01	0.009	29
TSO	B	1C-NT	0.032	0.03	0.016	34
TSO	B	1C-T	0.012	0.014	0.015	34
TSO	B	3C-NT	0.024	0.0385	0.014	24
TSO	B	3C-T	0.013	0.0185	0.015	24
TSO	V	1C-NT	0.014	0.015	0.01	34
TSO	V	1C-T	0.009	0.009	0.011	34
TSO	V	3C-NT	0.013	0.023	0.008	24
TSO	V	3C-T	0.007	0.0115	0.009	24
TSO	I_c	1C-NT	0.021	0.017	0.017	34
TSO	I_c	1C-T	0.011	0.012	0.018	34
TSO	I_c	3C-NT	0.013	0.0175	0.015	24
TSO	I_c	3C-T	0.009	0.0135	0.016	24

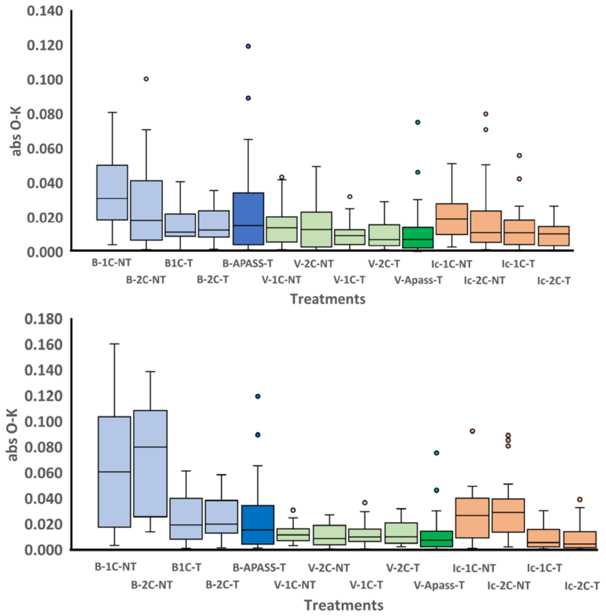

Figure 2. Boxplots of different treatments of absolute O–K values for different treatments Landolt standard stars taken with the telescopes and CCD cameras at (upper plot) the Tigh Speuran Observatory (TSO) and (lower plot) The Live Oak Observatory (LOO), with a comparison of APASS photometry on selected Landolt fields. Median values are horizonal bars within the quartile variation boxes, ranges are the vertical bars, outliers are circles.

and untransformed treatment/analyses (0.006–0.017) were almost identical.

Given the general trend that untransformed data appear less accurate (i.e., larger absolute O–K value), on the whole, than transformed data, we evaluated the null hypothesis that absolute O–K medians were statistically equal in paired treatments (Table 3). In general, tests between transformed versus untransformed absolute O–K values were significantly different (e.g., 1C-NT versus 1C-T). In contrast, tests between transformed data (1C-T versus 3C-T or 1C-NT versus 3C-NT) were not significant at the p = 0.05 significance level (Table 3). There were three exceptions, two Cousins I_c from TSO and one Johnson V from LOO (easily identified in Figure 2).

Given the precision values in Table 2, we evaluated the null hypothesis that there was a correlation between precision and accuracy as measured by absolute O–K values (Table 4). We found only five of 24 correlations to be significant (p = 0.05). We note that the stars used in this study were picked by us based on what we interpreted as stars with sufficient SNR (e.g., SNR > 20) to expect reasonable photometry. We conclude that under these conditions there is no correlation between accuracy and precision. This supports the fact that accuracy and precision measure different uncertainties (i.e., systematic error vs. random error). We would expect a correlation between SNR, magnitude, and precision (with standard deviation increasing as SNR decreases), but that was not studied here, nor could we investigate the relationship between SNR and accuracy because LesvePhotometry does not return SNR values for our target/check stars. We speculate that pushing the limits of target SNR under difficult or suboptimal seeing conditions would affect both precision and accuracy.

Given Wilcoxon analyses (Table 3) that showed a predominance of significant differences between the accuracy of transformed and untransformed analyses, we explored the effect of transformation on color correlation by performing least-squares fits to pairs of comparable treatments with the color index B–V as the independent variable and the "raw" untransformed O–K value as the dependent variable. The null hypothesis for each analysis is that the slope fit is not statistically different from a slope = 0 (p = 0.05). Acceptance of the null hypothesis (and its associated small coefficient of determination) is interpreted herein as a successful color transformation as there would be no statistical association between the B–V color index of a star and its estimated magnitude for a particular filter. Rejection of the null (and a higher coefficient of determination) would imply either untransformed or poorly transformed estimates.

Regression analyses results are reported in Table 4. All untransformed analyses reject the null hypotheses. That is, in all untransformed analysis there was a significant slope fitted to the data and the magnitude of that slope was significantly different from the null of slope = 0. In contrast, twenty-one transformed analyses accept the null. That is, the line fitted to the data have a slope that is statistically flat (slope = 0). A visualization of two regression analyses from TSO for the Johnson B filter using a single comparison star and the same ensemble using three comparison stars are shown in Figure 3 to illustrate these differences. The transformed fits are close

Table 3. Wilcoxon test results, CCD[1].

Observatory	*Filter*	*Treatment*	*N*	*z-Value p*
TSO	B	1C-NT/1C_T	33	<0.001*
TSO	B	3C-NT/3C-T	24	0.006*
TSO	B	1C-T/3C-T	24	0 .920
TSO	B	1C-NT/3C-NT	24	0.406
TSO	V	1C-NT/1C_T	24	0 .004*
TSO	V	3C-NT/3C-T	23	0 .031*
TSO	V	1C-T/3C-T	22	0.162
TSO	V	1C-NT/3C-NT	20	0.379
TSO	I_c	1C-NT/1C_T	22	0.072
TSO	I_c	3C-NT/3C-T	23	0.11
TSO	I_c	1C-T/3C-T	24	0.575
TSO	I_c	1C-NT/3C-NT	21	0.453
LOO	B	1C-NT/1C-T	24	0.001*
LOO	B	2C-NT /2C-T	24	0.001*
LOO	B	1C-T/2C-T	22	0.952
LOO	B	1C-T/allC-T	23	0.412
LOO	B	1C-NT/2C-NT	24	0.646
LOO	I_c	1C-NT/1C-T	23	0.001*
LOO	I_c	2C-NT/2C-T	17	0.01*
LOO	I_c	1C-T/2C-T	22	0.952
LOO	I_c	1C-T/allC-T	19	0.184
LOO	I_c	1C-NT/2C-NT	24	0.124
LOO	V	1C-NT/1C-T	19	0.276
LOO	V	2C-NT /2C-T	18	0.003*
LOO	V	1C-T/2C-T	24	0.944
LOO	V	1C-T/allC-T	24	0.834
LOO	V	1C-NT/2C-NT	24	0.124

[1] *N, number of observations varies due to ties.*

Table 4. Correlation and regression analyses of CCD Observations[1].

Obs.	*Filter/ Treatment*	*Corr. Pearson r*	*Regress p-value*	*R-sqr.*	*p-value*	*N*
LOO	B/1C-NT	0.1064	0.61	0.8	< 0.001*	25
LOO	B/1C-T	0.197	0.36	0.153	0.059	24
LOO	B/2C-NT	0.3726	0.04*	0.65	< 0.0004*	31
LOO	B/2C-T	0.1303	0.48	0.022	0.423	31
LOO	V/1C_NT	0.042	0.85	0.216	0.022*	23
LOO	V/1C-T	0.0899	0.68	0.338	0.003*	24
LOO	V/2C-NT	0.0407	0.83	0.065	0.167	30
LOO	V/2C-T	0.0793	0.68	0.259	0.004*	30
LOO	I_c/1C-NT	0.2099	0.32	0.671	< 0.00001*	24
LOO	I_c/1C-T	0.4171	0.04*	0.025	0.769	24
LOO	I_c/2C-NT	0.0605	0.75	0.15	0.034*	30
LOO	I_c/2C-T	0.3349	0.08	0.011	0.589	24
TSO	B/1C-NT	0.2489	0.16	0.513	< 0.001*	33
TSO	B/1C-T	0.2874	0.1	0.033	0.315	32
TSO	B/3C-NT	0.4021	0.046*	0.697	< 0.001*	33
TSO	B/3C-T	0.4181	0.053	0.006	0.723	25
TSO	V/1C_NT	0.0038	0.98	0.448	< 0.001*	32
TSO	V/1C-T	0.1843	0.3	0.126	0.042*	33
TSO	V/3C-NT	0.0108	0.96	0.409	0.001*	24
TSO	V/3C-T	0.2041	0.35	0.086	0.173	22
TSO	I_c/1C-NT	0.2655	0.13	0.413	< 0.001*	33
TSO	I_c/1C-T	0.794	<0.001*	0.004	0.741	33
TSO	I_c/3C-NT	0.3059	0.15	0.277	0.007*	33
TSO	I_c/3C-T	0.737	<0.001*	0.054	0.286	23

[1] *Filter/Treatment is filter and treatment; Corr. Pearson r is Pearson r of the correlation between the absolute O–K and the standard deviation of N stars; p(r) probability of rejecting the null hypothesis that absolute O–K values are correlated with standard deviation (a measure of precision). R-sqr. is the coefficient of determination of (B–V|untransformed O–K) of N stars; p(R-sq) tests the null hypothesis is that the slope of the least squares fit is zero (0). N is the number of Landolt standard stars used in each analysis.*

Table 5. Summary data, combined standard fields, CMOS.

Observatory	*Filter*	*Treatment*	*Median abs. O–K*	*IQR O–K*	*Precision*	*N*
BSM-NH2	B	1C-NT	0.006	0.039	0.004	18
BSM-NH2	B	1C-T	0.016	0.016	0.004	18
BSM-NH2	B	3C-NT	0.016	0.016	0.01	14
BSM-NH2	B	3C-T	0.011	0.012	0.01	14
BSM-NH2	V	1C-NT	0.016	0.022	0.004	18
BSM-NH2	V	1C-T	0.022	0.027	0.003	18
BSM-NH2	V	3C-NT	0.009	0.017	0.01	14
BSM-NH2	V	3C-T	0.012	0.016	0.01	14
BSM-NH2	I_c	1C-NT	0.019	0.032	0.008	18
BSM-NH2	I_c	1C-T	0.015	0.023	0.009	18
BSM-NH2	I_c	3C-NT	0.014	0.025	0.014	14
BSM-NH2	I_c	3C-T	0.014	0.011	0.015	14

Table 6. Wilcoxon test results, CMOS and CMOS/CCD[1].

Observatory	*Filter*	*Treatment*	*N*	*z-Value p*	*w-Value p*
BSM-NH2	B	1C-NT/1C-T	18	0. 347	>0.05
BSM-NH2	B	3C-NT/3C-T	14	†0.022*	<0.5
BSM-NH2	V	1C-NT/1C-T	18	0.928	>0.05
BSM-NH2	V	3C-NT/3C-T	14	0.726	>0.05
BSM-NH2	I_c	1C-NT/1C-T	18	0.267	>0.05
BSM-NH2	I_c	3C-NT/3C-T	14	0.952	>0.05
CMOS/CCD	B*	1C-NT	18	0.040*	<0.05
CMOS/CCD	B*	1C-T	18	0.031*	<0.05
CMOS/CCD	B	3C-NT	14	0.529	>0.05
CMOS/CCD	B	3C-T	14	0.298	>0.05
CMOS/CCD	V	1C-NT	15	0.177	>0.05
CMOS/CCD	V**	1C-T	18	0.025*	<0.05
CMOS/CCD	V	3C-NT	11	0.424	>0.05
CMOS/CCD	V	3C-T	11	0.424	>0.05
CMOS/CCD	I_c	1C-NT	18	0.447	>0.05
CMOS/CCD	I_c	1C-T	18	0.171	>0.05
CMOS/CCD	I_c	3C-NT	14	0.826	>0.05
CMOS/CCD	I_c	3C-T	13	0.384	>0.05

[1] BSM-NH2 are COMS-to-COMS tests, CMOS/CCD tests medians obtained for the same sample with different sensors. An asterisk in the z-value column marks rejection of the null hypothesis that the medians are equal. A single asterisk (*) in the Filter column denotes that CMOS absolute O–K estimates were significantly more accurate than CCD estimates; a double asterisk (**) denotes CCD estimate are more accurate.

Table 7.Correlation and regression analyses of CMOS observations[1].

Obs.	*Fltr/Treat*	*r*	*p(r)*	*R-sqr.*	*p(R-sqr.)*	*N*
	Abs. O–K | StDev			*B–V | O–K*		
BSM-NH2	B/1C-NT	0.139	0.584	0.408	0.004*	18
BSM-NH2	B/1C-T	0.571	0.013*	0.008	0.724	18
BSM-NH2	V/1C NT	0.339	0.169	0.291	0.021*	18
BSM-NH2	V/1C-T	0.231	0.356	0.025	0.531	18
BSM-NH2	I_c/1C-NT	0.432	0.073	0.089	0.28	18
BSM-NH2	I_c/1C-T	0.66	0.003*	0.014	0.646	18
BSM-NH2	B/3C-NT	0.154	0.598	0.517	0.004*	14
BSM-NH2	B/3C-T	0.703	0.005*	0.0004	0.996	14
BSM-NH2	V/3C NT	0.433	0.122	0.339	0.029*	14
BSM-NH2	V/3C-T	0.307	0.307	0.138	0.221	14
BSM-NH2	I_c/3C-NT	0.544	0.044*	0.055	0.418	14
BSM-NH2	I_c/3C-T	0.74	0.002*	0.073	0.352	14

[1] Fltr/Treat is filter and treatment; r is Pearson r of the correlation between the absolute O–K and the standard deviation of N stars; p(r) probability of rejecting the null hypothesis that absolute O–K values are correlated with standard deviation (a measure of precision). R-sqr. is the coefficient of determination of (B–V | "raw" O–K) of N stars; p(R-sq) tests the null hypothesis that the slope of the least squares fit is zero (0). N is the number of Landolt standard stars used in each analysis.

to zero slope and the scatter of actual magnitude estimates are less than the untransformed fits. This demonstrates visually that there is no relationship between the color of the star (B–V) and the magnitude, thus the transformed data are successfully color-transformed. In contrast the "raw" untransformed O–K estimates show a significant slope, as expected given that they are not transformed.

5. CMOS precision and accuracy—results and discussion

CMOS analyses are similar to CCD analyses except that only two Landolt Fields were used. Table 5 documents median values and variation for absolute O–K values as well as average precision as estimated from standard deviations of ten individual measures. Figure 4 visualizes these data. Note that the overall variation is less than the CCD data and the effects of transformation are less. Transformation yields consistently better results in accuracy in only three of the six pairs of treatments and a tie in one treatment pair (3C-NT versus 3C-T). The Johnson V-filter median results with untransformed data are more accurate than the median transformed results. However, when we look at the variation as measured by the IQR scores we observe that five of six pairs show less variation in the transformed results, suggesting that on the whole more accurate star magnitudes are achieved by transforming the data.

The Wilcoxon signed values pair-wise test results for the CMOS data are quite different from the CCD results (Table 6), reflecting the slight differences in medians shown in Table 5. We found only a single test result (i.e., 3C-NT versus 3C-T) to be significant.

Least-squares analyses are reported in Table 7. In spite of the failure of the Wilcoxon tests to favor one treatment over another (with one exception), the least squares fits do demonstrate why we have pointed earlier to variation around the median values shown in Table 5. The null hypothesis (slope = 0) is rejected in all untransformed regression analyses but only in two of the regression analyses of transformed data. We interpret this to mean that more of the stars measured had improved color transformed magnitude estimates compared to their estimates in untransformed treatments, in spite of the fact that the medians are similar. That is, transformation decreased the scatter and shifted the scatter towards the y = 0 axis. We conclude that transformation is, in fact, effective in increasing accuracy in these data.

We also examined precision. We found precision uncorrelated with accuracy for this sample of stars (Table 7), a result similar to the CCD analyses.

Wilcoxon signed-value pair-wise tests were used to evaluate the null hypothesis that accuracy, as measured by absolute O–K values, of similar CCD and CMOS magnitude estimates were statistically similar between treatments ($p = 0.05$). The results, using two fields imaged at TSO and BSM-NH2 (Table 6, lower), show that nine of the twelve pair-wise tests were insignificant.

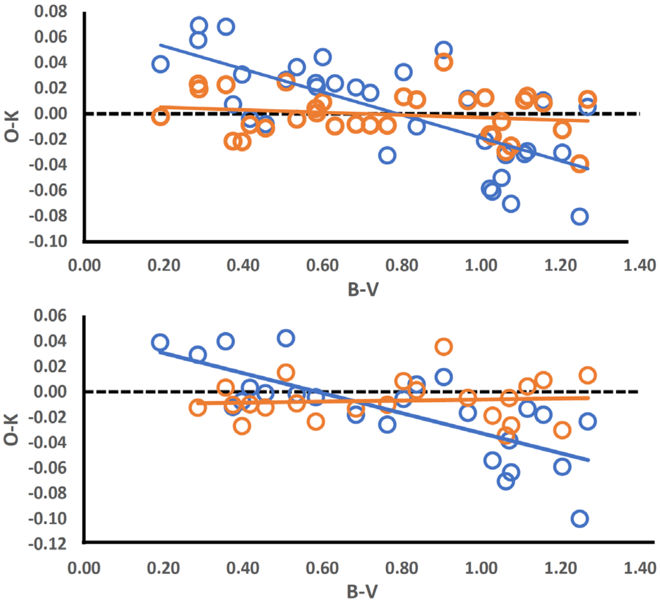

Figure 3. Regression analyses of color (B–V) and raw OK values (O–K) of transformed and untransformed analysis of Landolt standard stars for the Johnson B filter taken at the Tigh Speuran Observatory (TSO). In each case blue circles are data points and blue lines are least squares fit to data of non-transformed data while orange circles are data points, and the orange line is the least squares fit of transformed data. (Upper plot): Analyses using one comparison star (1C-NT blue, 1C-T orange). (Lower plot): Analyses using three comparison stars (3C-NT blue, 3C-T orange).

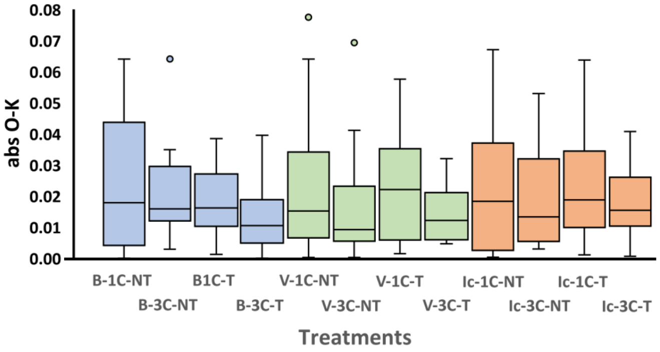

Figure 4. Boxplots of different treatments of absolute O–K values for different treatments of Landolt standard stars taken with the BSM-NH2 telescope and CMOS camera. Median values are horizonal bars with the quartile variation boxes, ranges are the vertical bars, outliers are circles.

Two of the tests (Johnson B, 1C-NT, and 1C-T) were significant with CMOS data being more accurate than CCD data. One test (Johnson V, 1C-T) was significant with CCD data being more accurate.

6. Conclusions and recommendations

Our study suggests a procedure for amateur photometrists to measure and assess the accuracy of their photometry and improve amateur photometry. Specifically, we propose that one should image Landolt Standard Fields repeatedly (N = 10) and assess Observed–Known (O–K) magnitudes for many standard stars. This procedure confirms that transformation significantly improves the accuracy of measured target magnitudes. Our results suggest that an average accuracy between ±0.02 magnitude is achievable without extraordinary efforts to pick comparisons stars of the same magnitude and color but with adequate SNR and good calibration. No doubt close attention to issues such as transformation coefficients, calibration, and comparison star choice leads to further improvement after an initial assessment.

We identified seven conclusions from our efforts:

1. Transformation improves accuracy, especially in cases where one is forced to pick a comparison star that is different in color than the target. We did not evaluate accuracy without transformation in situations where a comparison star and target are of similar color due to the nature of the study, but it can be improved by picking a comp star as close in color as possible to the target star.

2. There are no statistical differences between single-comp and ensemble comp methods shown in this study. However, we prefer ensemble methods because they can result in statistically meaningful measurement uncertainties given three or more comps using the standard deviation of all the comparison stars. We did not evaluate whether more stars in an ensemble than used in this study would result in greater accuracy than sole use of a single comparison star.

3. CCD and CMOS cameras were equally accurate in estimating magnitude.

4. Simple tests such as the Wilcoxon tests for similarities to median values may not provide a definitive answer to the effect of transformation. Least-squares fits provide a view of the entire data set and are more definitive.

5. Accuracy and precision are uncorrelated given adequate signal to noise ratios of targets and comparison stars. Precision (random error) uncertainty is not a measure of accuracy (systematic error). That said, there are many research programs for which precision is of upmost importance and accuracy is of secondary importance.

6. For variables with long periods compared to the imaging cadence, the most direct way to measure precision is to compute the mean and standard deviation of the magnitudes of a short time series of 4-10 images and report the mean as the calculated magnitude and standard deviation as the uncertainty (precision).

7. Directly measuring the accuracy of a variable is not possible. One can, of course, measure the fit of the observation to a model (O–C analysis), but this is different from accuracy as used in this study. Rather, one can do O–K analysis on the check star(s) that stand as secondary standards in the analysis or rely on the standard deviation of the ensemble variation.

7. Acknowledgements

This study would not have been possible without the resources of the American Association of Variable Star Observers, including catalogs of Standard Star photometry and charts, the AAVSO Bright Star Monitor Network, and the APASS catalog (Henden *et al.* 2018). Our thanks to Pierre de Ponthière for his help with LesvePhotometry. We thank the reviewer and editor for insightful comments and the editorial staff for their hard work.

References

de Ponthière, P. 2011, LesvePhotometry, automatic photometry software (http://www.dppobservatory.net).

Henden, A. A., Templeton, M., Terrell, D., Smith, T. C., Levine, S., and Welch, D. 2018, AAVSO Photometric All-Sky Survey, data release 10 (https://www.aavso.org/apass).

Koppelmann, M. 2005. in *Proceedings for the 24th Annual Conference of the Society for Astronomical Sciences*, eds. B. D. Warner, D. Mais, D. A. Kenyon, J. Foote, Society for Astronomical Sciences, Rancho Cucamonga, CA, 107.

Landolt, A. U. 1983, *Astron. J.*, **88**, 439.

Landolt, A. U. 2013, *Astron. J.*, **146**, 131 (doi:10.1088/0004-6256/146/5/131).

Mandel, J. 1964, *The Statistical Analysis of Experimental Data*, Dover Publ., Inc. New York.

Newberry, M. V. 1999, in *Precision CCD Photometry*, ASP Conf. Ser. 189, eds. E. R. Crain, D. L. Crawford, R. A. Tucker, Astronomical Society of the Pacific, San Francisco, CA, 74.

Sonnett, S., Meech, K., Jedicke, R., Bus, S., Tonry, J., and Hainaut, O. 2013, *Publ. Astron. Soc. Pacific*, **125**, 456.

Stubbs, C. W., and Tonry, J. L. 2006, *Astrophys. J.*, **646**, 1436.