Draft version April 16, 2018
Typeset using LaTeX **modern** style in AASTeX62

A SYSTEMATIC APPROACH TO COMBINING DATA FROM MULTIPLE OBSERVERS APPLIED TO COMET HALE-BOPP'S VISUAL LIGHTCURVE

Lead author: M. Womack[1]

[1] *University of South Florida*
Department of Physics
4202 E. Fowler Ave
Tampa, FL, 33620 USA

1. MOTIVATION

In analyzing visual data from multiple observers, the questions inevitably arise of which data to reject, and under what justification, and whether combining data from observers, each with his or her own systematic errors, leads to a biased result. Without instrumental calibration, there is no certain answer to these questions, but as discussed by (Kidger 2002; Mousis et al. 2014), such calibration is itself problematic, and in any case is not available for the observations discussed here.

We offer a systematic approach to combining data from multiple observers yielding a self-consistent consensus fit. In application to comet C/1995 O1 (Hale-Bopp), the procedure does not significantly affect the grossest measure, namely the slope of magnitude *vs.* log distance, but, applied to data already corrected to heliocentric distances and for phase, does reduce the *statistical* error bars.

We assume three categories of errors:

1. Every observer reports the brightness of an object on a scale that is shifted up or down from other observers, but by the same number of magnitudes, $\delta_{\rm obs}$, independent of distance or brightness. Without instrumental calibration, we can best estimate $\delta_{\rm obs}$ as that observer's mean deviation from a consensus fit to the data (that is, an average).

2. Some observers may have a slope bias, underestimating the brightness of dimmer objects and overestimating those of brighter ones, or *vice-versa*. While it is difficult to correct for such error without calibration, the bias can be detected (relative to the consensus fit), and that observer's data discarded.

3. Finally, some observers may have a great deal of scatter in their data but no bias. We can weight these observations less in fits.

1.1. *Method*

We seek a consensus fit to the data. The comet exhibits small but noticeable deviations from a power law (intensity $\sim r^{-s}$) on time (distance) scales larger than any outbursts. In particular, nearly all pre-perihelion data between 2 and 1 a.u. suggest a negative curvature in the graph of magnitude against log distance.

A straight-line fit would penalize observations that report this feature accurately. Based on the total number of apparent features between the closest (0.91 a.u.) and furthest (9.4 a.u. post-perihelion) observation, we fit to fifth-order polynomials. The effect is to smooth the data.

The deviations between an observer's measurements and the consensus fit at the same distance are considered noise. The set of all such deviations by one observer defines the noise distribution for that observer, characterized by a mean δ_{obs}, variance σ^2_{obs}, skewness, excess kurtosis, *etc.*

To compute the self-consistent fit, we iterate the following steps until convergence to a fractional tolerance of 0.0001 (absolute tolerance if any fitting parameter is less than 0.0001) of all six polynomial fitting parameters:

1. We fit a fifth-order polynomial through all the data of the (possibly shifted) magnitudes against log distance by the method of least squares (Legendre 1806), weighting each observation inversely as the observer's variance of deviations, σ^2_{obs} Initially, the magnitudes have been corrected for geocentric distance and for phase angle, but have not been shifted. Also initially, the weights are all equal, because we do not know the distributions of deviations.

2. Using the same data as in 1, we can also fit a straight line, recovering the slope (power law) $-s$ and statistical error bars on that fit based on the shifted, weighted data.

3. For each observer, we consider the distribution of deviations between the observations and the fit at the given distance (*i.e.*, in the graph, the vertical vectors between data points and the polynomial fit). For each observer separately, we treat the distribution as noise and estimate the mean, δ_{obs}, and variance, σ^2_{obs}.

4. For each observer separately, we shift all magnitudes by $-\delta_{\text{obs}}$ as computed in Step 3.

For the present data set, these iterations converge to the specified tolerance after between ten and twelve iterations. We now have a self-consistent fit and set of shifted data. Row 4 of Table 1 shows small changes in best-fit slopes after the procedure. This is reflected in visibly smaller scatter in the lightcurve; the statistical error fitting to a straight line has been cut roughly in half. Note that non-straight-line features in the consensus polynomial fit limit how far the statistical error bars in a straight-line fit can shrink.

Table 1. Best-fit slopes for pre-perihelion and post-perihelion obervations before and after corrections and the self-consistent method of this appendix. Lines 4 and 5 use self-consistently shifted magnitudes and weight each observer's data inversely as σ_{obs}, the variance of the noise. The fitting method reduces statistical error but does not significantly alter the slope estimates.

	pre-perihelion		post-perihelion	
1. raw data	11.33	±0.08	12.66	±0.07
2. geocentric distance correction only	8.06	±0.05	9.14	±0.07
3. geocentric distance and phase corrections	9.14	±0.05	9.92	±0.07
4. self-consistent shifts	9.15	±0.02	9.85	±0.03
5. drop observers, self-consistent shifts	9.09	±0.03	9.80	±0.04

The self-consistent procedure has eliminated the need to discard data arbitrarily (*e.g.*, points more than some number of standard deviations above or below a consensus fit, which otherwise would throw off least-squares fits) by weighting points inversely as the observer's variance. However, as noted above, an observer whose systematic error changes with brightness would still affect the slope adversely. We can detect such a systematic problem by applying Student's t-test to each observer's data set, comparing the difference of the mean deviation from the consensus fit in the first half of the observer's data (sorted by distance) to the mean deviation in the second half. Since the variances of the two halves may not be equal, we normalize by the "pooled variance" to get an approximate t-statistic. Assuming approximately Gaussian noise (not always justified), we calcluate the p (probability) value that the t-statistic would be as large as observed or larger under the null hypothesis that first and second halves of the data were drawn from the same distribution, *i.e.*, that the observer did not contribute bias to the slope relative to the consensus.

For a p value less than 0.05, we reject the null hypothesis and say that the observer "fails" the t-test; that is, his or her data bias the slope. We then discard such observers from the data set and repeat the self-consistent iteration (steps 1–4). If repeating the t-test at the end of second set of iterations results in more observers failing the t-test, we then repeat the procedure until no new observers are discarded. For the pre-perihelion data, one new observer failed the t-test after the second set of iterations, so we iterated a third time without that observer's data. At the end of the third set of iterations, no new observers failed the t-test.

As shown by Line 5 of Table 1.1, throwing out observers results in small changes in the slopes. For the pre-perihelion data, seven observers out of 17 were discarded, accounting for 430 of the original 1,003 data points. For post-perihelion data, four out of 12 observers were discarded, accounting for 232 out of the original 486 data points.

For comet Hale-Bopp, the self-consistent procedure resulted in adjustments to the slope roughly comparable to the original statistical error bars, while cutting these approximately in half. These changes are far smaller than those associated with correcting for heliocentric distance and phase. The observations considered in the present work were taken by skilled amateurs calibrated by eye to stars in generally reliable catalogs. Future work may rely on less homogeneous amateur networks, in which case a self-consistent method for combining and weighting magnitudes, and discarding subsets with systematic slope bias, could prove useful.

REFERENCES

"Spanish Monitoring of Comets: Making Sense of Amateur Photometric Data," EM&P, 90, 259-268.

Legendre, A.M., "Nouvelles Méthodes pour la détermination des Orbites des Cométes," Courcier, VIII, 1-55, 4, DCCC. f. 10.

Mousis, O. and 40 co-authors, "Instrumental methods for professional and amateur collaborations in planetary astronomy," Experimental Astronomy, 38, 91-191.