FROM ARCHIVES TO ANALYSIS: COMPUTERS REDEFINE THE AAVSO MISSION

Grant Foster
AAVSO
25 Birch Street
Cambridge, MA 02138

Received: October 26, 1992

Abstract

Recent advances in computer technology and program development promise great progress for the AAVSO in the near future. Not only have we improved our methods of archiving and evaluating variable star observations, but we will also soon be ready to enter the field of substantial in-house data analysis.

Over the last five years, thanks to several grants from NASA's Astrophysics Division, the AAVSO has modernized its computer hardware with the acquisition of IBM personal computers, now tied together with sophisticated network software. Networking of any kind requires powerful computers, and our host systems are exactly that, with a potent new 486 processor. Further, all our systems now boast EGA monitors, giving each workstation the high-resolution graphics capability that makes working with everything from Fourier transforms to 3-D plots of chaotic attractors to documents for the *Journal* much cleaner, more precise, and just plain fun.

In addition to changes in system software and hardware, we have continued apace our primary purpose: the archiving, evaluation, and dissemination of variable star observations. Last year brought in another 276,210 observations from around the world, which were digitized and added to the archives. We also made substantial progress towards completing the digitization of data prior to 1963. Finally, the data from *AAVSO Reports 28, 29, 30,* and *31* were merged into the archives, so that now we have at our fingertips all observations from 1963 to the present. Our computerized database contains over five million observations.

All this has to do with the day-to-day operation of our primary mission. Yet clearly the state of the art in astronomy, instrumentation, computers, and mathematical analysis continues to march inexorably forward; despite our intense focus on maintaining our database, we must also continue to explore new techniques for old goals, and begin to define new goals for ourselves. Thus it is especially fortunate that the AAVSO leadership is tolerant of unorthodox ideas and schedules, and encourages new explorations.

Two years ago I developed two experimental programs: EDIT and GFT. EDIT is an on-screen editing and evaluation program, designed to give the operator a live light-curve plot (not unlike the light curves we print with our computer), color-coded to indicate the status of the various data. The user can then move about with the cursor to locate, identify, and tag various data points. In its initial form, EDIT was crude but marvelously effective. What would have taken many steps and several days could now be accomplished with a single procedure, in minutes. Because of its popularity with Headquarters' technical staff, it not only gained acceptance, it also underwent revisions due to a flood of requests for improvements. This was the beginning of a most serendipitous trend; in fact it is the teamwork here at Headquarters which anchors all our program development, especially for new concepts like EDIT. EDIT is now able automatically to locate a star's data in our archives, or in any other data file. The user can search for a particular observer or set of observers, to look for trends in one person's data or to use as "standard candles" to define the overall behavior of the star. Most important (although invisible to the user), EDIT now processes files in binary mode, allowing extremely rapid data acquisition and updates.

GFT, the other experiment, is a sophisticated Fourier-transform program. It implements the data-compensated discrete Fourier transform, or DCDFT (Ferraz-Mello 1981), an algorithm I learned about from a program written at the Maria Mitchell Observatory (Belserene 1986). Fourier transforms are the subject of many dissertations; suffice it to say the DCDFT compensates for difficulties that arise when observations are made at irregular time intervals (like most astronomical data). I added nothing mathematically new to the algorithm, but I did enable the user to have more direct interaction with the analysis and more immediate graphic display of the results and data. I also expanded the data-handling capacity of the program to take advantage of newer, larger personal computers.

About a year ago we took the next logical step: the marriage of the two programs. I conceived the idea to combine in one software package the data viewing and manipulation capacities of EDIT with the sophisticated mathematical analysis of GFT; the result is ZAP, our do-it-all data evaluation and analysis program. At first, ZAP was written for my own numerical analyses; due to its immediate popularity, it quickly burgeoned into much more.

ZAP now incorporates (at least in part) three important elements. First, it provides fast access to all AAVSO data. Stars can be identified by designation or by name, and data can be retrieved either from a user-specified file or from the AAVSO archives. Files are processed in binary mode for fast access. Furthermore, ZAP can process files either in standard AAVSO format, or in simple columnar format, with the user specifying which columns contain the relevant quantities (time and magnitude). ZAP can currently accommodate 10,000 data points in memory. This large limit enables the user to load all available data on a star, or a large time interval for those stars more densely observed. Once the data are loaded, the user can view the entire time interval, focus on a sub-interval, or move from time frame to time frame, progressively stepping through the data in high detail. Figure 1 shows a typical ZAP screen, displaying a light curve of unevaluated AAVSO data on Mira for the thousand days from JD 2447000 to 2448000.

Second, ZAP makes it easy to view and evaluate the data. In evaluation, no data point is ever deleted from the AAVSO archives, but it may be tagged so as to be excluded from analysis. The user controls the viewing parameters, choosing the display limits both for JD and for magnitude. The data are color-coded, and the user controls the cursor, identifying observations and (if so desired) tagging them with edit markers. ZAP even includes a "zoom-in" window, so the user, when exploring a densely crowded region of data, can zero in on a particular point. By connecting a given observer's observations, the user can track someone's work against the general flow of data or edit someone's observations of a star wholesale, and the edit markers can be preserved in the archives, in real time. ZAP even provides immediate access to the actual file record representing a given observation. Figure 2 shows the same Mira data, with lines connecting four observers' individual data.

All through the process of evaluation and exploration, ZAP offers the user repeated opportunities to save measurements, notes, and computations to disk files for later analysis, allowing ZAP to be much more than just an editor. Any and all results can be preserved, so the interesting surprise that may emerge from a ZAP session need not be lost or forgotten.

Third, the system has at its disposal (and will have even more) powerful, sophisticated mathematical tools. As previously noted, ZAP was conceived as the union of EDIT and GFT; in its initial incarnation it incorporated the potent data-compensated discrete Fourier transform. But not all mathematical tools are so esoteric - in fact, the most useful and most used are the simplest. Therefore we can compute simple means (one of the most common statistics in variable star analysis), including computation of variances, enabling the program, upon request, to display an error envelope (representing a two-standard-deviation plus-or-minus range) as well as smaller error bars indicating the uncertainty present in the estimated average itself. The error envelope is a useful guide (but not a panacea) for estimating the tolerable range of variation. Figure 3 illustrates a display of averages, with an error envelope, for a ninety-five day interval around one of the maxima of Mira.

A bit more complex, but with advantages of its own, is to fit a curve by polynomial regression. ZAP allows the user to fit (by the method of least squares) a polynomial to the selected time frame. The user can choose the polynomial degree, from a simple first-order straight line (to reveal linear trends), to a twentieth-order curve. Again, the results of a polynomial fit can be saved for later analysis.

These straightforward techniques accomplish what I consider to be one of the most necessary and effective steps for the analysis of heavily observed stars: reduction of the number of data points. This can be accomplished by averaging, fitting polynomials, and many other methods. For our most heavily-observed program stars (which include some of the most interesting) we have a wealth of data. To load and track all of a star's data can be a computationally daunting assignment. A span of a mere thousand days can tax the limits of the most powerful computers. When the data are reduced, the sheer number of data points is dramatically decreased, with little or no sacrifice in information content. The lesser number of data points can then be put through the mathematical wringer, without spending enormous processing time. Also, the time spacing of the reduced data is at the user's control, enabling the generation of an equally-spaced time series, although if gaps are present in the original data they cannot be filled in. Such a series is naturally suited to the fastest Fourier transform algorithms, and is the basis for the most straightforward construction of what is called a "phase space" suitable for chaos analysis.

Other period-search algorithms include Phase-Dispersion-Minimization, or PDM (Stellingwerf 1978), and autocorrelation (Burki, Maeder, and Rufener 1978; Percy, Jakate, and Matthews 1981; Baliunas *et al.* 1983). PDM tests a trial period by using it to assign phases to the data, then plotting phase versus magnitude and computing the dispersion in the data. If the data are periodic, the true period will align points of equal phase, and the dispersion will be minimized. From a purely mathematical viewpoint, PDM is not unlike Fourier analysis. Both of them match the data to periodic trial functions with various degrees of freedom. However, PDM has three advantages: first, for a given trial period the scan includes as many degrees of freedom as the user chooses "bins" in the phase plot (while Fourier analysis has two degrees of freedom, amplitude and phase, per trial period); second, PDM makes no assumption about the shape of the light curve, only its periodicity (Fourier methods match the data to sinusoidal curves); and third, PDM does not require the evaluation of transcendental functions (sine and cosine).

Autocorrelation is an old method which, in my opinion, has been inadequately appreciated. It looks for similarity between the light curve and a copy of the light curve which is offset by a time lag. It too has the advantages of multiple degrees of freedom for a given trial, and of insensitivity to the shape of the light curve. Furthermore, it can detect *characteristic* time scales, even in data which are nowhere near periodic. Its usefulness has been amply demonstrated in recent research.

What will be the eventual destiny of this program, which began its life to satisfy the idle curiosity of an idle programmer? First of all, we will complete the incorporation of mathematical algorithms, so that ZAP will come to be a standard tool for all kinds of analysis in addition to exploring and evaluating data. Second, we have a lot of work to do to document the program, an effort which has only barely begun. Eventually, we hope to make ZAP available to other users. As yet the distribution of this and other AAVSO software is undecided, but we are considering making it available to our members, and using a simplified version as the basis of a data analysis software package to be part of our educational initiative, Hands-on Astrophysics: Variable Stars in the Physics / Math Lab. We need to be conservative, even uncompromising, about releasing a quality product; hence it will be some time before any version is ready for distribution. If we are equal to the task, we may perform a great service to the cause of astronomy, and give some of today's students a glimpse of just how much fun it will be to join the ranks of tomorrow's astronomers.

Whatever the future of ZAP, this much cannot be denied: as computers grow more powerful, as mathematical tools become more discriminating, as the demands of data analysis grow ever more pressing, the AAVSO must enter uncharted territory with energy and skill, so that as astronomy enters the 21st century, the AAVSO can continue to be one of its most productive participants. The technology has so exploded that we are limited only by the power of our imagination. Yet amidst the excitement of new techniques and new missions for the AAVSO, we must never compromise our primary mission, the duty we have fulfilled for eighty years, and will continue for the future: to make the observations of amateur astronomers available to professional astronomers and all those involved in research. We pledge not to lose sight of the fact that the AAVSO is here for *you*, our members, and *you*, the astronomers, and thank goodness that it is all we can manage just to keep up with you.

[Ed. note: since this paper was presented in September 1992, changes in computer hardware and software have continued to advance, with the generous support of grants from NASA and other sources. We promise further updates in articles to come.]

References

Baliunas, S. L. *et al.* 1983, *Astrophys. J.*, 275, 752.
Belserene, E. P. 1986, in *The Study of Variable Stars Using Small Telescopes,* ed. John R. Percy, Cambridge University Press.
Burki, G., Maeder, A., and Rufener, F. 1978, *Astron. Astrophys.*, 65, 363.
Ferraz-Mello, S. 1981, *Astron. J.*, 86, 619.
Percy, J. R., Jakate, S. M., and Matthews, J. M. 1981, *Astron. J.*, 86, 53.
Stellingwerf, R. F. 1978, *Astrophys. J.*, 224, 953.

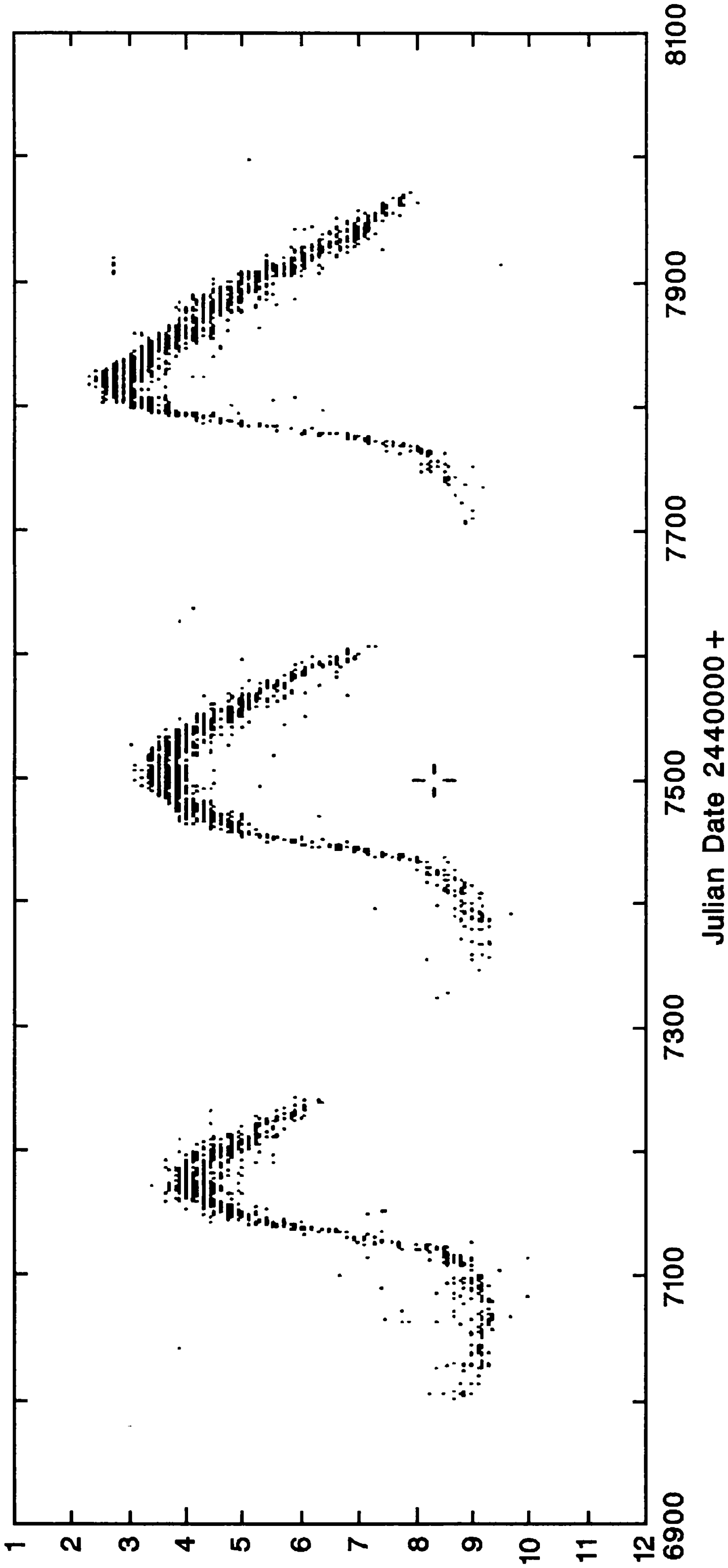

Figure 1. A typical ZAP screen, displaying an unevaluated light curve of Mira for JD 2447000 - 2448000.

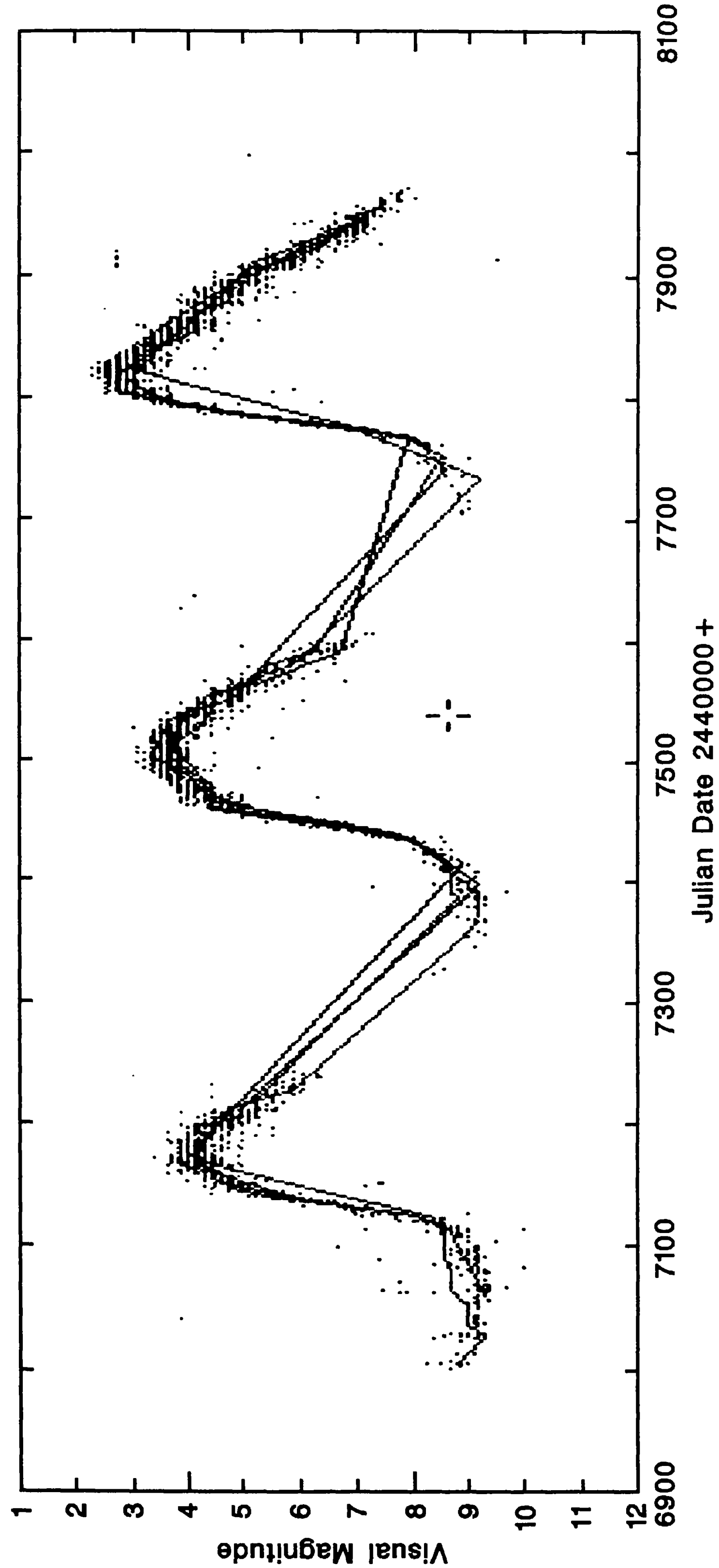

Figure 2. The same Mira data as in Figure 1, with lines connecting four observers' individual data.

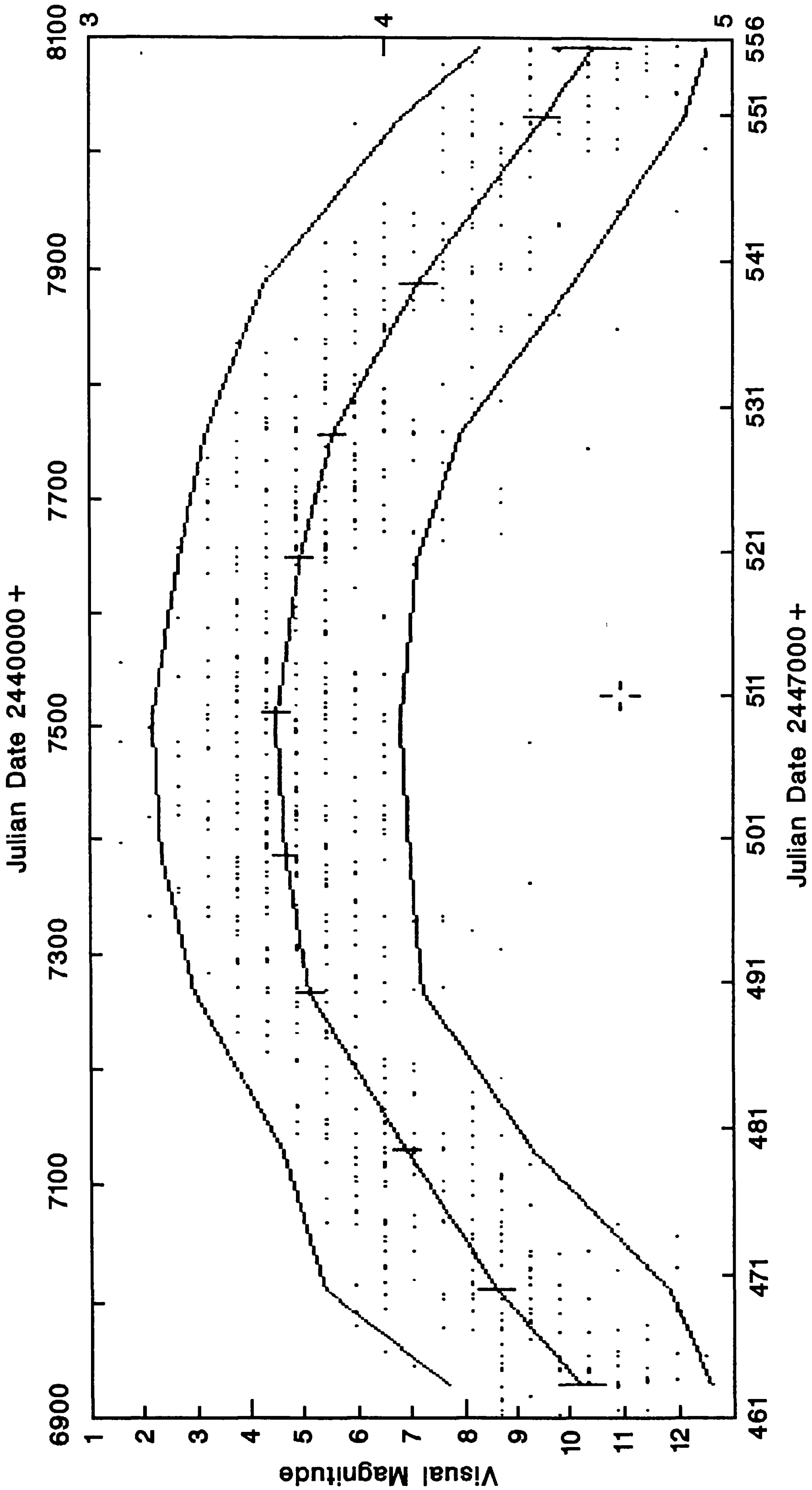

Figure 3. A ZAP screen showing a maximum of Mira with ten-day means (middle line), error bars on the means, and a two-sigma error envelope. Points outside the two-sigma envelope have been tagged as discrepant and so were excluded from the analysis.