

On the Automatic Determination of Light-Curve Parameters for Cepheid Variables

PETER B. STETSON

Dominion Astrophysical Observatory, Herzberg Institute of Astrophysics, National Research Council, 5071 West Saanich Road,
Victoria, British Columbia V8X 4M6, Canada

Electronic mail: peter.stetson@hia.nrc.ca

Received 1996 March 28; accepted 1996 June 27

ABSTRACT. A computerized algorithm for the automatic detection of Cepheid variables and for the estimation of their periods, amplitudes, and mean magnitudes from sparse datasets is presented. It is intended to be suitable for use in such programs as the measurement of Cepheid distances to external galaxies, for example, with the *Hubble Space Telescope*. The reliability of the algorithm is tested by application to new photometric reductions of prerepair *HST* images of the nearby Sdm galaxy IC 4182, with comparison to published analyses of the same data (Saha, et al. 1994, ApJ, 425, 14).

1. INTRODUCTION

The launch and subsequent repair of the *Hubble Space Telescope* (*HST*) have led to a surge of interest in Cepheid variable stars as a means to calibrate the extragalactic distance scale. Several teams are employing either *HST* or ground-based telescopes to detect Cepheids in external galaxies, determine their periods and mean magnitudes, and infer the galaxies' distances by comparison of their apparent period–brightness relation with the absolute period–luminosity relation measured in the Galaxy and the Magellanic Clouds (e.g., Sandage et al. 1992; Saha et al. 1994; Freedman et al. 1994; Pierce et al. 1994; Tanvir et al. 1995).

Modern detectors, both on *HST* and on the ground, obtain millions of pixels' worth of data in a single brief exposure, and each observation can record tens of thousands of stars. To be sure, the photographic plates of fond memory can contain even more resolution elements than current CCDs, but such data are acquired in analog form and are most easily searched for variable stars by visual examination (“blinking”). (The comparatively low quantum efficiency of plates also helps to keep the data rate down to manageable levels.) Electronic images are more usually acquired and stored in digital form, and while they can be displayed and blinked as analog images—and it is instructive to do so—they are also well suited to analysis by more rapid, more impersonal, and less tedious automatic techniques.

The “more rapid” and “less tedious” aspects of automatic data analysis speak for themselves, but the “more impersonal” part deserves some expansion. In discovering and measuring a sample of variable stars—such as Cepheids for the purpose of distance determinations—several different types of confidence in the results are desired. First, we wish to be confident that the star detected as a “variable” is a physical variable, and not merely displaying a time sequence of random noise or image contamination. Second, we wish to be confident that the type of variability is correctly identified, e.g., that the star is in fact a classical Cepheid variable as opposed to a Mira or an eclipsing binary. Third, we wish to

be confident that the relevant parameters (viz., the period and mean magnitude in the case of a Cepheid) are correctly determined within known tolerances. And last, we wish to be confident that our sample is as complete as possible and—more important—is either unbiased with respect to the important parameters, or subject to bias whose nature is quantitatively known and can be allowed for. In the case of Cepheid distances, the *sine qua non* is a sample which, for some known range of period, fairly samples all luminosities that are possible at those periods. Biases in parameters of secondary importance, e.g., a bias against variables of low amplitude, can be tolerated to the extent that those secondary parameters are uncorrelated with luminosity, or to the extent that an identical bias operates in the reference sample of Galactic and Magellanic Cloud Cepheids.

The human brain includes a wonderful image processor, and an experienced investigator can do a remarkable job of identifying true, physical variable stars under conditions of varying background and differing seeing. Then the expert astronomer can judge the correctness of a given trial period by the apparent “rightness” of the phased light curve. The personal biases of any given investigator are hard to define quantitatively, however, and the samples of variables that would be obtained from the same data by investigators of different experience levels or predispositions might be sufficiently dissimilar as to compromise the science goals of the study. It is not so much a question of whether different astronomers will discover different variables, but whether they will assign the same confidence level to the correctness of the classifications and periods: in other words, whether individual subjectivity will lead them to accept and reject the same dubious data points in the absence of some universally recognized, objective confidence criteria. A properly designed automatic algorithm could have its biases adequately mapped out by comparison with the discovery efficiency of one or more expert astronomers and also through the medium of tests on artificial variable stars digitally inserted into images of a star field. Then this algorithm could be used with equal facility, reliability, and repeatability by investigators of

all experience levels. By operating directly on measured magnitudes (whether instrumental or calibrated) and their associated standard errors, an automatic digital algorithm can assign quantitative confidence levels to the various stages in the detection, classification, and measurement of a variable star. Without the dynamic-range limitations of the human eye and—most important—by considering simultaneously the whole corpus of observations of a given star rather than by a pairwise comparison of the images, the sensitivity of the algorithm to low-amplitude variables can be maximized.

Welch and Stetson (1993; hereafter, WS) presented a simple algorithm for identifying highly probable variable stars from lists of (magnitude, error) data for stars observed numerous times. The method relied on a measurement of the time-dependent correlation of magnitude *residuals*: given a mean magnitude determined from a string of observations, do observations taken close together in time show differences from that mean that are consistent to a statistically significant level? By scaling the residuals according to the quantitatively estimated standard error of each magnitude determination, the algorithm provides a numerical confidence index for the physical reality of the detection that is independent of magnitude, although, of course, the larger the random errors of the individual magnitude measurements, the larger the amplitude of physical variation required to achieve a given confidence level. Furthermore, an appropriate choice of the correlation statistic makes the method quite robust against the contamination and blunders inherent to real data (as distinguished from statistical idealizations): in particular, individual incorrect measurements tend to *reduce* the magnitude of the WS numerical index, unlike some more venerable statistical indices such as overall magnitude variance. The WS technique is especially appropriate when the data are naturally paired as, for instance, when images were obtained in two photometric bandpasses or in cosmic-ray-split pairs each time the field was revisited.

The present paper attempts to extend the general concept of impersonal digital methodology to the identification of Cepheid variables, in particular, and to the determination of their relevant physical parameters: period, phase, amplitude, and mean magnitude. It must be stressed that software “black boxes” will *not* supplant experienced astronomers in the foreseeable future. The technique described here represents something much more modest than an attempt “to invent and implement a perfect procedure for the automatic detection and classification of variable stars” (anonymous referee). Rather, it represents an *initial* prototype with much shorter-term goals: (a) to serve as an additional, independent and objective, method of searching for variable stars to supplement (not, anytime soon, replace) currently accepted techniques; and (b) to help provide some insight into just where the essential information content of a string of photometric data lies. How can variable detection be optimized? When changing a particular criterion or parameter, does the method work better or worse? Can we understand why? How do the experts do what they do, and how can their skills be most effectively explained to novices? In the referee’s words, “The obvious power of independent analysis of a dataset by different scientists will vanish if we were to sim-

ply and solely rely on the output produced by a software package.”

In the present prototype, detection of variable candidates is based on a variant of the WS algorithm—it has been necessary to de-sophisticate it slightly to allow for the case where the data are not always taken in pairs, and to make it still more robust against the possibility of multiple corrupt observations. Then the recognition of a particular candidate variable as a candidate *Cepheid* and the subsequent estimation of its light-curve parameters is done through the technique of template matching, since it has been shown that the shape of a Cepheid light curve is a nearly unique function of its period. Thus, the modified WS index provides the principal measure of confidence that the variability is real, and not due to noise. The quality of the fit of the data to the template then provides a quantitative measure of confidence that the light curve is that of a normal classical Cepheid; the appropriateness of the derived period, amplitude, magnitude, and color can provide secondary indicators of the correctness of the classification. Then, the availability of completely sampled template light curves optimizes the determination of flux-weighted mean magnitudes from datasets that may be sparse, may be poorly sampled in phase, and may consist of observations with significantly different standard errors. Finally, the quality of the template fit also provides quantitative confidence intervals for the derived period and mean magnitude. Thus, each Cepheid that appears in a given target’s period–magnitude relation can be given its own individual, numerical weight reflecting the probability that the variation is real, the likelihood that the classification as a Cepheid is correct, and the degree of confidence in the measured values of the two relevant quantities. These properties make the proposed algorithm particularly promising for the analysis of Cepheid data from *HST*, where the scarcity of the resource encourages the smallest possible number of visits to the target; where the need to orient the solar panels limits the length of the observing season at any given telescope orientation; where a cosmic-ray event or detector blemish can corrupt the measurement of a variable candidate at any given epoch; where safing events can disrupt the most carefully optimized observing schedule; and where the need to work near the detection limit yields magnitude standard errors that change significantly from maximum to minimum light. At present, due to the immediate desire to cope with data from *HST*, the algorithm has been calibrated specifically to the photometric bandpasses Johnson *V* and Kron–Cousins *I*. Should future experience show the method to be efficacious and convenient, it can easily be extended to other bandpasses.

The use of the new algorithm will be illustrated by application to new reductions of the prerepair *HST* images of the nearby Sdm galaxy IC 4182, which was host to the type Ia supernova 1937C. These observations were obtained in 1992 January through March and in 1993 January, during the course of program No. 2547, “Calibration of Supernovae of Type Ia as Standard Candles” (A. Sandage, principal investigator). The 1992 data have previously been discussed by Sandage et al. (1992) and Saha et al. (1994), who employed a battery of both semiautomatic and personal techniques to

obtain and cross-check their variable candidate lists. The data from the 1993 revisit have not previously been studied, to the best of my knowledge. The extraction of instrumental magnitudes from the original images was done with ALL-FRAME (Stetson 1994), using the same methodology and model point-spread functions as were used by Freedman et al. (1994).

2. IDENTIFYING CANDIDATE VARIABLES

The Welch/Stetson variability index I is defined by the equation

$$I = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{b_i - \bar{b}}{\sigma_{b,i}} \right) \left(\frac{v_i - \bar{v}}{\sigma_{v,i}} \right)},$$

where b_i and v_i are the apparent magnitudes obtained for the candidate star in two observations closely spaced in time on some occasion i , $\sigma_{b,i}$ and $\sigma_{v,i}$ are the standard errors of those magnitudes, \bar{b} and \bar{v} are the weighted mean magnitudes in the two filters, and n is the number of observation pairs. This notation is rooted in the assumption that on each visit to the program field a single pair of observations in the B and V photometric bandpasses is obtained. The method would work equally well if two observations in the *same* bandpass were obtained at each visit—as in cosmic-ray split pairs obtained with *HST*. In this case, $\bar{b} = \bar{v}$ may be taken to be the mean of all $2n$ observations, rather than the means of two different samples each of size n . WS showed that this formulation is both effective in identifying variables, and robust against a small number of corrupt observations: a single observation with an inordinately large residual actually tends to drive the index toward a slightly *smaller* value (i.e., makes the star *less* likely to be flagged as a variable), unless of course a second blunder of the same sign is made in the other observation of that pair.

In images obtained with WF/PC or WFPC2 on *HST*, however, stellar images damaged by cosmic rays, warm pixels, or other blemishes are sufficiently common that it seems worthwhile to modify the WS index to make it still more robust against defective data. At the same time, it is necessary to generalize the formulation to allow the inclusion of data from visits where only a single frame was obtained (not all *HST* data are taken as cosmic-ray splits) or where more than two observations were made in close succession. A more robust version of the index is

$$J = \frac{\sum_{k=1}^n w_k \operatorname{sgn}(P_k) \sqrt{|P_k|}}{\sum_{k=1}^n w_k}, \quad (1)$$

where the user has defined k pairs of observations to be considered, each with a weight w_k ; $P_k = \delta_{i(k)} \delta_{j(k)}$ is the product of the normalized residuals of the two observations, i and j , constituting the k th pair; and δ is the magnitude residual of a given observation from the average of all observations in that same bandpass scaled by the standard error ($= (v - \bar{v})/\sigma_v$, for instance); I adopt this modified notation to remove the implication that the observations in a pair will necessarily be in different filters. In fact, it is not quite optimal to define δ to be the residual of one magnitude measurement from the mean of all measurements in that filter. Be-

cause that observation itself has been used in defining the mean of all observations, the residual from the sample mean is expected to be smaller, on average, than the residual of an observation from “truth” by a factor that depends upon the number of observations in the sample. Any basic statistics text will show that the size of this bias is $\sqrt{(n-1)/n}$, where n is the total number of observations contributing to the mean. Since a given frame pair may include data from two filters which did not have equal numbers of observations overall, I can therefore define the “relative error”

$$\delta = \sqrt{\frac{n}{n-1}} \frac{v - \bar{v}}{\sigma_v},$$

and so on, thus allowing all residuals to be compared on an equal basis. This is the definition of δ which will be used in all that follows.

On a visit to the program field where a single pair of frames was obtained, $i(k)$ and $j(k)$ will naturally point to the two magnitudes derived from those two observations; the weight w_k of such an observation pair may be taken to be unity. In a dataset consisting of magnitude pairs with purely random noise, the expectation value $\langle \delta_i \delta_j \rangle$ is zero; thus J should tend to zero for a nonvariable star, and will tend toward some positive number for a physical variable. However, if only a single exposure was obtained on some occasion, we could set $i(k) = j(k)$, contributing the square of the relative error of that one observation to the summation. In the limiting case where *only* single observations were obtained at the different epochs, J would then reduce to a robust measure of the standard deviation of the observed magnitudes relative to that expected from the errors of the individual observations: a measure of the external repeatability relative to the internal precision. In a random dataset containing purely single observations with known standard errors, the expectation value $\langle \delta_i^2 \rangle$ is unity. It is therefore necessary to subtract this offset to render the expectation value for single observations containing only random noise equal to zero, the same as for paired observations;

$$P_k = \begin{cases} \delta_{i(k)} \delta_{j(k)}, & \text{if } i(k) \neq j(k); \\ \delta_{i(k)}^2 - 1, & \text{if } i(k) = j(k). \end{cases}$$

Otherwise, a given quantitative value of J would mean very different things for stars contained in different numbers of paired and single exposures. The weights w_k of single observations may be taken to be unity if all observations are equally valid, because the ratio of external repeatability to internal precision is itself an efficient index of variability in the absence of corrupt data. However, if corrupt observations are common, it may be desirable to assign reduced weight to a singleton observation, because the square of a botched magnitude will necessarily make a large positive contribution to the index.

Conversely, if more than two observations are all obtained within a time span \ll the shortest periodicity being sought, individual observations may be included in more than one pair. For instance, three closely spaced observations abc may be distributed into the three pairs ab , bc , and ac , with two-thirds weight being assigned to each pair so

that in the aggregate they contribute double weight (as a variance determined from three observations carries twice the weight of a variance estimated from only two).

As one step in the robustification of the new index, the average magnitude used here is no longer the simple weighted arithmetic mean of all observations in a given bandpass; it is a robust mean based on the numerical gimmick discussed before by Stetson (1987, Sec. III D 2 *d*; and 1989, Chap. 3 B): after an initial determination of the weighted mean has been made, stars are reweighted according to the size of their residuals. The stars' weights are multiplied by a factor

$$\left[1 + \left(\frac{|\delta|}{a}\right)^b\right]^{-1},$$

and the mean is redetermined with the new weights. The procedure is iterated until the mean and the individual weights stabilize. The same technique is also used to render the template light-curve fits discussed below robust against corrupt observations. I prefer this approach to the much more widely adopted " $k\sigma$ clipping" for several reasons. Most important, with small datasets $k\sigma$ clipping can hold an entire dataset hostage to a single observation: if the rejection threshold is set at 3σ , then clipping will accept an observation with a 2.99σ residual, and accord it unit weight. Conversely, it will reject an observation with a 3.01σ residual, giving it zero weight. Thus, an insignificant change in a single datum—and that datum already somewhat suspect to begin with—can make a significant change in the final answer. This is an undesirable property in an algorithm. By smoothly tapering the weights from unity for $\delta=0$ to zero for $\delta\rightarrow\infty$ one ensures that a small change in any one datum will always lead to a tiny change in the answer. Second, clipping can make the ultimate iterated solution to a nonlinear problem dependent on the input starting guesses for the fitting parameters: if the initial model identifies a particular observation as erroneous, the observation is ignored and may never influence the final result, whereas if the initial model happened to pass near that point, it is given unit weight and is allowed to attract the solution still closer. With continuous weights, even an outlying observation exerts some minor attraction on the solution, and the final answer is achieved by balancing the weighted pulls exerted by all the data. I do not believe that this approach can be justified analytically from first statistical principles—indeed, unless the true form of the error distribution (including small, Gaussian error contributors and large, non-Gaussian blunder contributors) is known in quantitative detail, no algorithm can be proven to be optimal. The most I can say is that this approach has certain philosophically desirable properties, and appears to work as well as can be expected under the circumstances. For my present purposes, I have adopted $a=b=2$. Extensive experimentation with artificial corrupt datasets (*op. cit.*) indicates that the efficacy of the method is not very sensitive to these quantities. Note that the square root inside the upper summation of Eq. (1) renders the index still more robust by reducing the impact of extreme outliers relative to that of more typical residuals.

A value for J may be quickly computed for every star measured in an ensemble of images; those with the largest values of J are the most probable variables. However, when the number of independent observations is small and the possibility of corrupt data is non-negligible, it is useful to have some backup measure of the way the observed magnitudes are distributed between the maximum and minimum values. For instance, if a light curve were a pure noiseless sawtooth, then all magnitudes between the maximum and minimum would be equally likely; if the light curve were a pure sinusoid, values near the extrema would be more probable than values near the mean; for a constant star subject to Gaussian errors, magnitudes near the mean would be more likely than near the extrema; and for a constant star with one corrupt observation, values near the robust mean would be very common, and there would be a single remote outlier. I therefore take as my backup index a robust measure of the kurtosis of the magnitude histogram:

$$K = \frac{1/N \sum_{i=1}^N |\delta_i|}{\sqrt{1/N \sum_{i=1}^N \delta_i^2}}, \quad (2)$$

where in this case the index i runs over all N observations available for the star without regard to pairing. It may easily be shown that, in the limit where the total range of variation is vastly larger than the σ 's of the individual observations, for a hypothetical square-wave light curve $K=1.0$; for a pure sinusoid $K\rightarrow\sqrt{8/\pi}=0.900$; for a sawtooth (uniform distribution of magnitude probabilities) $K\rightarrow\sqrt{12}/4=0.866$; for a Gaussian magnitude distribution $K\rightarrow\sqrt{2/\pi}=0.798$; and for a single measuring blunder of size Δ , $K=(|\Delta|/N)/\sqrt{\Delta^2/N}=1/\sqrt{N}\rightarrow 0$ as $N\rightarrow\infty$. In the other extreme, where individual random measuring errors dominate over the physical variation, the Gaussian limit is approached; $K\rightarrow 0.798$. We can therefore make our algorithm more capable of distinguishing desirable types of variation from undesirable ones by letting our provisional index be $JK/0.798$; this will be numerically equal to J when the nature of the perceived variation is Gaussian (e.g., the observational errors have been underestimated, for instance, or a star varies erratically), but it will be amplified by a small factor for typical smoothly varying light curves, and suppressed by a large factor for extreme but infrequent discrepant magnitudes (e.g., some data are corrupt, or the candidate is a widely detached eclipsing binary).

Finally, a purely practical consideration: in determining the distance to a galaxy from Cepheid light curves, we are most interested in stars with the largest possible number of observations. When a star is absent from one or more images of a given field, this is likely to indicate a problem: the star is near the edge of the frame or near a detector blemish, so minor pointing changes cause it to be lost sometimes; or it is very near the detection limit. Therefore, we can multiply our variability index by a further factor of $\Sigma w/w_{\text{all}}$, where w_{all} is the total weight a star would have if successfully measured in all frame pairs. This way, when a set of candidate variable stars is selected from the corpus of data corresponding to a given target field, those candidates that were successfully measured the most times will be the *first* to be followed up.

The final variability index proposed here is thus

$$L = \left(\frac{JK}{0.798} \right) \left(\frac{\sum w}{w_{\text{all}}} \right). \quad (3)$$

A value for L can be determined for every star in the program field having some minimum total weight ($\sum w \geq w_{\text{min}}$), and stars exceeding some threshold value of L may be subjected to period searches and light-curve fits.

3. FITTING THE TEMPLATE LIGHT CURVES

Before the template Cepheid light curves (see the Appendix) can be fit to the data for a given star by iterative linearized least squares, it is necessary to come up with initial guesses for the period and phase of the variation, because the equations are nonlinear in these quantities. I adopt a period-finding algorithm which is a variant of the Lafler–Kinman (1965) string-length technique. Their principal formula for evaluating the quality of any given trial period was

$$\Theta = \frac{\sum_{i=1}^N (m_i - m_{i+1})^2}{\sum_{i=1}^N (m_i - \bar{m})^2},$$

where the N individual magnitude observations m have been sorted in order of increasing phase according to the trial period, and for phase closure $m_{N+1} \equiv m_1$. The numerator of the fraction is a measure of the length of “string” that would be required to connect the points in order of phase, while the denominator is a standard measure of the width of the magnitude histogram without regard to phasing. The “best” period for some candidate variable is that for which Θ is minimized, and if that minimum value of Θ is sufficiently small compared to unity, the physical reality of the variation may be taken as likely. The Lafler–Kinman index—unlike some alternative methods, such as certain implementations of phase-dispersion minimization (e.g., Stellingwerf 1978)—does not employ any binning of the data, but rather treats each observation on its own. This may be an advantage for many modern studies, where the number of independent epochs may be small (of order 10–20).

However, some adjustments to the Lafler–Kinman formula may be envisioned to make it more robust against a small admixture of corrupt data. The first of these is straightforward: replace the squares of the magnitude differences in the summations with the absolute values of those differences. This has the usual effect of suppressing the importance of the most extreme outliers relative to more “typical” differences. Second, it is necessary to make allowance for the fact that the standard errors of the individual observations may differ. Unlike the classical studies where variable stars were investigated with photomultipliers or photographic plates and exposure times could be chosen to make the signal-to-noise ratio of the measurement comparable at all phases, in the type of work contemplated here some of the variable candidates may be near the detection limit, so the standard error of the magnitude is a strong function of phase. Clearly, a large difference between two poorly determined magnitudes is of less consequence than the same difference between two well-determined ones. The answer is to weight the magnitude differences by the standard errors of the constituent magnitudes.

Finally, it is necessary to note that with a small number of observations for any given variable candidate, it is not possible to insure that the entire cycle is fairly sampled—for any given trial period there may be sizeable gaps of phase with no observations. Therefore I adopt a scheme where the weight of a given term in the sum is multiplied by a quantity $\sim (\phi_{i+1} - \phi_i)^{-1}$, where $\phi = (t_i - t_1)/P$, under the assumption that a large magnitude difference between observations that are closely spaced in phase is important evidence that the trial period is wrong (and therefore should make a large contribution to the string length) whereas a large magnitude difference between observations at widely separated phases is not as strong evidence that the postulated period is incorrect. I therefore adopt the formulation

$$S(P) = \frac{\sum_{i=1}^N w(i, i+1) |m_i - m_{i+1}|}{\sum_{i=1}^N w(i, i+1)}, \quad (4)$$

where

$$w(i, i+1) = \left[\frac{1}{\sigma_i^2 + \sigma_{i+1}^2} \right] \left[\frac{1}{(\phi_{i+1} - \phi_i) + \epsilon} \right].$$

An identical formula has been independently developed by Freedman and Madore (in preparation). The softening parameter ϵ is required to keep two observations that happen to fall at the same phase from having infinite weight. It seems adequate to take $\epsilon = 1/N$, so if identical-phase observations have unit weight, two observations separated by the average phase spacing will have half-weight. Because the purpose here is solely to sort out which possible periods are of the greatest interest for a given variable candidate, the normalization factor represented by the denominator of the Lafler–Kinman ratio is omitted here; the role of a numerical index quantifying the likelihood of the physical reality of the proposed variation is taken instead by the index L defined above. Thus, for various trial periods, values of S are calculated, and up to 20 of the deepest local minima of the function $S(P_j)$ are retained in a data array, for $S(P_j)$ less than some lintel (opposite of “threshold”) value. In my current implementation, the minimum period to be tested is specified by the user; it was found in early trials that when the number of epochs sampled is small (e.g., ≤ 20), then testing periods that were too short (compared to that given, for instance, by the Nyquist frequency corresponding to the shortest interval between successive revisits) resulted in testing very many possible permutations of the data points. This frequently led to spuriously attractive light curves, just from the possibility of accidentally arranging observations in order of a progression of random magnitude errors. Therefore the astronomer is required to exert some modicum of physical sense: in particular, not searching for variables with periods less than X days in galaxies where such Cepheids should be far too faint to detect. This will result in missing some other types of short-period variables, such as eclipsing binaries, but this is a loss that can often be accepted. Starting with the minimum desired period, successive trial periods are chosen such that the change in phase of the last observation of the series relative to the first is equal to 0.02:

$$P_{j+1}^{-1} = P_j^{-1} - \frac{0.02}{\Delta t},$$

where Δt is the total amount of elapsed time between the first and last observation of the series; the maximum period considered is Δt or 100 days, whichever is less; and the *largest* local minimum value of the string length retained as a possible period is taken to be the mean value \bar{S} , averaged over all trial periods. After all periods corresponding to local minima in the string length meeting the above criteria have been located, they are sorted in order of increasing string length, and the periods corresponding to the 20 shortest strings are subjected to template light-curve fits.

Before the fits can proceed, however, for each of the possible trial periods P for a given variable candidate, it is also necessary to obtain an initial guess at the time of zero light-curve phase, since the fitting equations are nonlinear in this quantity, too. This is done in a simple and crude way by the solution of the *linear* least-squares problem,

$$m_i = m_0 + a \cos(\omega_i) + b \sin(\omega_i),$$

where for my present purpose $\omega_i = 2\pi\phi_i = 2\pi(t_i - t_1)/P$. This yields the (crudely) estimated epoch of minimum visual light in the fundamental harmonic as

$$t_0 = t_1 + P \tan^{-1}(b/a)/2\pi.$$

Clearly, this is not a correct estimate of the zero epoch for a light curve that is not a simple sinusoid. But provided careful note is taken of the signs of a and b , it does yield an estimate that at least is in the right quadrant, which is all that is required. The solution also yields starting guesses for the mean magnitude (m_0) and semiamplitude ($\sqrt{a^2 + b^2}$) of the light curve. It should be remembered that this least-squares fit and the subsequent, more sophisticated, fits of the light-curve templates are rendered robust against corrupt data by the iterative reweighting method outlined above.

Finally, the observed data for each candidate variable can be fitted to the family of template light curves (see the Appendix): for each trial period generated by the string-length algorithm, the V -band and I -band light curves of appropriate shape are computed from the template relations, and the solution is iterated to optimize the determinations of the five fundamental parameters: V_0 , I_0 , P , A , and t_0 . If observations in the two filters are comparable in terms of number and quality, they both contribute to the determination of period, fundamental-frequency visual semiamplitude, and epoch of zero phase; the relative importance of I compared to V is reduced only by the factor ~ 0.6 representing the ratio of the I -band amplitude relative to that in V . If—as is the case with many *HST* programs—observations in V greatly outnumber those in I , the former will dominate the choice of the suitable template, and the latter will primarily determine I_0 alone.

The calibration of the instrumental magnitudes to the standard system occurs simultaneously with the light-curve fits. Having read in the appropriate zero points and color transformations, the program uses the provisional light curves in V and I to predict the color that the star should have at each given epoch; the instrumental magnitude corre-

sponding to that instant of time is then transformed to the standard photometric system using the color transformation appropriate to that bandpass. When this has been done for all the observations, the residuals from the provisional V and I light curves are used to compute incremental corrections to the mean magnitudes, the period, the amplitude, and the epoch of minimum light in the fundamental harmonic of the Fourier expansion for V (my working definition of zero phase). The newly tweaked light curves are then used to recompute the predicted standard-system magnitudes and color at each epoch, and the whole process iterates to convergence. Thus, the actual predicted color of the variable at each epoch is used in the photometric transformation, not some mean color. When the solution has converged, the template light curves are used to predict the stars' magnitudes in V and I at each 0.01 of phase, these are converted to fluxes, and the flux-weighted mean magnitudes are computed from a simple trapezoidal-rule numerical integration. This method thus is minimally sensitive to a few spurious magnitudes, to a poor distribution of observations over phase, or to observations that are not simultaneous in the two photometric bandpasses.

Least-squares template fits are undertaken for all trial periods representing local minima in the string length, beginning with the starting guesses associated with the shortest string and working up toward the longest string-length retained. The choice of the best light curve from the up to 20 possibilities that were examined for each star is based on a simple scoring scheme. Features indicating a maximally desirable fit are: (1) the smallest possible standard error of unit weight after subtracting the template light curve from the phased data points; (2) the largest possible total weight in the solution (i.e., the smallest possible reduction in the weight of the observations by the residual-reweighting scheme alluded to above); and (3) a fitted amplitude most nearly appropriate to a Cepheid variable. As regards point (3), I originally considered that the *largest* possible fitted amplitude would be a reasonable indicator of the best light-curve fit, under the assumption that incorrectly phased observations should lead to a light curve with a small net amplitude and large scatter. This turned out to be a mistake. With a comparatively small number of observations, especially when the computer program was allowed to consider periods shorter than that given by the Nyquist frequency, it was very easy to find a trial period and time of minimum brightness that would concentrate the observational epochs near phases 0.25 and 0.75, with no good constraint on the actual values of maximum and minimum apparent magnitude near phases 0.0 and 0.5. Light curves of extremely large amplitude could be consistent with these datasets, even though the actual observed magnitudes showed no such range of variation. (In this context, "very easy" means that such spurious fits would occur a few times per thousand stars, which is comparable to or larger than the actual expected frequency of true Cepheids.) I therefore decided to adopt a scheme that would reject amplitudes that were unphysically large as well as uninterestingly small. Since the sample of Milky Way and Magellanic Cloud Cepheids which were used to define the template light curves were wholly contained within the inter-

val $0.194 \leq A \leq 0.587$, the score assigned to a given amplitude was

$$s = \{1 + [\log(A) + 0.46]/0.3\}^{-6}.$$

Thus, a trial light curve with an amplitude of $10^{-0.46} = 0.35$ mag receives a score of unity; amplitudes of 0.2 and 0.6 mag are both assigned a score of 0.80; a score of 0.5 is assigned when $|\log(A) - 0.46| = 0.3$ —that is, for $A = 0.17$ and 0.69 mag; and from there the score tapers smoothly and *rapidly* to 0.0 as $A \rightarrow 0$ or $A \rightarrow \infty$. The total score assigned to any fitted light curve is then

$$\text{score} = \left[\frac{m.e.1(\text{min})}{m.e.1} \right] \left[\frac{wt}{wt(\text{max})} \right] \left[\frac{s}{s(\text{max})} \right],$$

where the (min) and (max) values represent the minimum and maximum observed values of those statistics among the various provisional light-curve fits for a given star. A score of 1.0 indicates that a given template fit is the best one in all three categories, and is highly desirable; a highest observed score less than unity for a given star indicates that, while it may be the best overall fit, it is not the best according to all three criteria. The ratio of the score of the second-best fit to that of the best fit, which I denote $R_{2,1}$, is a quantitative indicator of the possibility that a false period has been accepted: one may assume this probability is 50–50 if the top two light curves have equal scores, and decreases to zero as this ratio goes to zero. This business of scoring the alternative light curves is the most highly experimental part of the whole automated procedure, and the most likely to be capable of substantial improvement.

As I undertook the reanalysis of the observations of IC 4182 that are described in the next section, I found that very often solutions originating from two different local minima of the string length would converge to the same solution. Furthermore, several different but very similar light curves could often be derived from the data for a given star, due to the possibility of different cycle counts across sizeable time gaps in the observational sequence, and sometimes due to the freedom to make different choices of possibly spurious observations for downweighting. Therefore I modified the program to make it not report a possible solution when an equal- or higher-score solution within 6% of the same period (0.025 in the base-ten logarithm) had already been found. Similarly, I had it not report a solution when another light curve within one standard error in magnitude, period, amplitude, and phase had an equal or higher score.

4. EXAMPLE: IC 4182

A field in the dwarf irregular galaxy IC 4182 was observed with the *Hubble Space Telescope* on 20 occasions between 1992 January 27 and 1992 March 13 (Sandage et al. 1992; Saha et al. 1994), and once again on 1993 January 24. On two of the 1992 visits and on the sole 1993 visit, a cosmic-ray split pair of V (F555W) and a pair of I (F785LP) images were obtained; on 17 of the 1992 visits a pair of V frames only was obtained; and on one of the 1992 visits the first planned V observation failed, resulting in a singleton observation at that epoch.

The photometric reduction of these data was performed with ALLFRAME (Stetson 1994), using the same point-spread functions for the various chip-filter combinations as were employed by Freedman et al. (1994) and Hughes et al. (1994) in their studies of WF/PC images of M81=NGC 3031. I employed essentially the same color calibration as was employed by Saha et al., except that for compatibility with my usual software packages, it was necessary to invert (approximately) their relations from

$$V = F555W - 0.0679(F555W - F785LP)$$

$$+ 0.0198(F555W - F785LP)^2,$$

$$I = F785LP + 0.0485(F555W - F785LP)$$

$$+ 0.0268(F555W - F785LP)^2$$

to

$$F555W = V + 0.077(V - I) - 0.023(V - I)^2,$$

$$F785LP = I - 0.055(V - I) - 0.030(V - I)^2.$$

I did not attempt to derive aperture corrections or fundamental zero points for my reductions; instead I worked directly from the PSF-fitting magnitudes and arbitrarily adjusted my zero points to yield approximate agreement, in the mean, with the Cepheid photometry of Saha et al. I have made no attempt to establish a true magnitude scale independently of that of Saha et al., and this paper should not be taken to represent a redetermination of the distance to IC 4182.

For the variable-star search I defined the image pairs in the obvious way, with same-visit same-filter observations being matched with each other, while the singleton V observation was paired with itself and given half-weight in comparison to the true-pair epochs. The immediate analysis is limited to just the 1992 data for the most meaningful comparisons with the prior work of Saha et al.; later, I will repeat parts of the present discussion including the 1993 visit. Figure 1 illustrates the values of the variability index L plotted against apparent V magnitude for all stars in the four WFC chips that were successfully measured in at least 16 of the 19 frame pairs from 1992. Fits to template light curves were attempted for all stars observed in at least 16 frame pairs and having values of L in excess of 0.9. String lengths were determined for trial periods ranging from 2.0 days to the length of the interval between the first and last observation, 46.2 days. During the robust, least-squares template fits, however, solutions were allowed to converge to values outside this range.

I remind the reader that the template light curves are not actually defined for periods less than 7 days (see the Appendix). Data for candidate variables with provisional periods shorter than this are actually fitted to simple sinusoids with equal amplitudes in V and I . This could be justified by the assertion that in more remote galaxies, variables with periods less than 7 days are likely to be comparable-mass eclipsing binaries, or that in nearby galaxies the shortest periods belong to overtone pulsators which tend to have more sinusoidal light curves. Really I adopted this stopgap due to the lack of any obvious alternative. While not providing a truly accurate representation of a Cepheid light curve, this approxima-

tion does nevertheless provide periods and mean magnitudes that are precise enough to be of interest. Similarly, the template light curves have not been defined for periods longer than 100 days. A variable with a candidate string-length period less than 100 days, but whose template-fitting solution converged to a period in excess of 100 days, was fitted with an extrapolated template. Similarly, a variable whose string-length period was greater than 7 days but which converged to a shorter period was likewise fitted with an extrapolated template, while a variable with a string-length period below 7 days which converged to a greater period continued to be fitted with a sinusoid. This approach was adopted for a practical reason: it is possible that for a given dataset a sinusoidal fit would converge to a period greater than 7 days, while the best fit of a Cepheid-like template to the same data would lie at a period less than 7 days. If the type of fit were to be switched every time the solution crossed the 7-day boundary, convergence would never be reached. The method adopted ensures that at least *some* reasonable fit will be achieved in a finite number of iterations.

There are two different ways in which subjective grades were applied to candidate variable stars. First, a visual examination of the phased light curve was made to judge whether it gives the impression of being “right”—if it did, it was assigned light-curve quality class 1. If there was some degree of doubt about the reality of the variation, or if the variability looked real but there was some suspicion that the fitted period was incorrect, it was assigned class 2. If the light curve seemed to be the result of poor data rather than physical stellar variability, it was assigned class 3. Second, an independent quality scale was based upon the appearance of the star in the median-average CCD image: a star having a peak clearly distinct from its neighbors was assigned image class 1; a star which looked real but was involved in the profile of a brighter star was assigned class 2; while a star which was hopelessly confused with multiple bright neighbors or was superimposed on a detector flaw was assigned class 3. Figure 2 is a duplicate of the L vs. V plot for only those stars that were considered as possible variables either by Saha et al. or by the present technique; here I have plotted the ordinate on a logarithmic scale to compress the top and expand the bottom of the diagram for better visibility. Symbol types are filled circles for possible and probable variables identified in Table 3 of Saha et al., and empty circles for additional stars identified by the present algorithm for which a visual examination of the fitted light curve gave me a subjective impression of being “right”; in both cases large symbols are used when I had assigned both light-curve quality class 1 and image quality class 1, while smaller symbols represent candidates that had scored at least one 2. All stars falling into these two categories are listed in Table 1. The \times 's in Fig. 2 represent stars which met the formal variability requirements of $n \geq 16$ and $L \geq 0.9$, but which scored at least one 3 in the subjective quality classifications. Among the 35 stars with $L > 1.7$, there are three with at least one 3 (9%), seven with at least one 2 (20%), and 27 with two 1's (77%); most of the 2's are highly probable long-period variables with periods greater than the span of the observations. Among the 35 stars with $1.1 < L < 1.7$ there are seventeen 3's

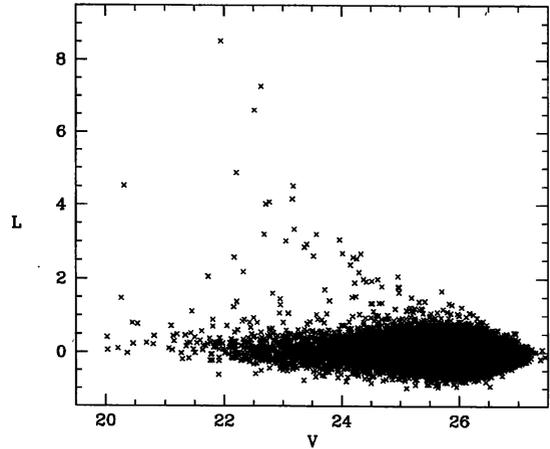


FIG. 1—Values of the variability index L , defined in the text, vs. apparent visual magnitude for all stars measured in all four chips for IC 4182.

(49%), eleven 2's (29%), and eight 1's (23%). Among the 38 stars with $0.9 < L < 1.1$ there are thirty 3's (79%), seven 2's (18%), and one 1 (3%).

The present method successfully recovered all but three of the 39 variable-star candidates identified by Saha et al. Two of these were their stars C3-V5 and C4-V3, for which they derived no period, and identified the type as “?” and “Red Var.?” respectively; on my scale these stars had $L = 0.410$ and 0.719 , and thus fell into the regime where it is difficult to regard the variation as statistically significant. The other star missed by the present method was C4-V1, identified as “Cepheid” by Saha et al. with a period of 3.67 days—my use of a large symbol to represent this star in Fig. 2 repre-

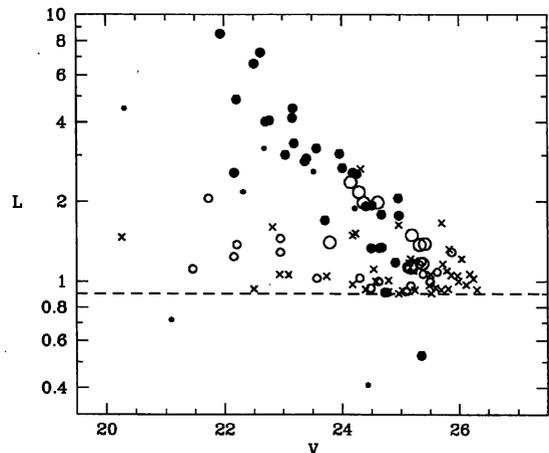


FIG. 2—Values of the variability index L , defined in the text, vs. apparent visual magnitude for all stars considered to be possible variables either according to the present study or that of Saha et al. (1994). Filled circles represent candidate variables identified by Saha et al., while empty circles represent candidates identified here for the first time. In each case, larger symbols represent variables with a high degree of subjective confidence, while smaller symbols are those with a lower degree of confidence. Crosses represent stars which were flagged as possible variables, but were rejected on the basis of a subjective judgment that the variation was probably spurious or that the stellar image was badly contaminated by crowding.

TABLE 1
Cepheid Parameters Derived from 43 Observations in 1992

Chip	x	y	ID	P	$\langle V \rangle$	$\langle I \rangle$	L	A	P	σ	$\langle V \rangle$	σ	$\langle I \rangle$	σ	$R_{2,1}$	P_2	Q
1	49	62	6	42.0	21.87	21.19	8.510	0.46	40.4	0.90	21.95	0.01	21.43	0.02	0.854	21.3	1
1	314	78	4	24.7	23.29	22.45	4.506	0.40	24.7	0.28	23.21	0.01	22.60	0.04	0.195	13.3	1
1	330	67	5	9.20	24.10	23.55	3.045	0.39	9.12	0.06	23.96	0.02	23.63	0.07	0.358	42.5	1
1	576	454	2.370	0.28	9.89	0.09	24.15	0.01	23.23	0.05	0.450	18.8	1
1	572	181	1.984	0.43	5.80	0.04	24.62	0.03	23.83	0.10	0.601	11.5	1
1	358	493	2	7.30	24.54	23.75	1.789	0.43	7.35	0.04	24.55	0.02	23.77	0.13	0.505	17.4	1
1	337	504	1.493	0.30	5.23	0.05	25.20	0.03	24.76	0.21	0.564	14.8	1
1	249	517	1	6.95	24.47	23.71	1.336	0.23	6.96	0.08	24.49	0.02	23.87	0.10	0.401	42.3	1
1	336	377	3	1.329	0.21	34.8	1.25	24.49	0.02	24.04	0.14	0.942	30.2	2
1	334	658	1.290	0.44	2.74	0.02	25.80	0.05	25.70	0.72	0.863	3.07	2
1	592	260	1.288	0.11	29.5	0.97	22.90	0.01	22.74	0.07	0.866	18.9	2
1	393	297	1.190	0.44	3.09	0.02	25.27	0.05	24.73	0.28	0.997	10.9	2
1	355	188	1.158	0.43	4.49	0.05	25.12	0.03	24.16	0.14	0.648	3.80	2
1	350	578	1.084	0.51	3.03	0.02	25.68	0.05	25.15	0.30	0.775	11.8	2
2	412	190	2	37.5	22.86	21.85	7.260	0.32	35.3	0.85	22.69	0.01	21.92	0.04	0.409	14.2	1
2	245	109	3	6.88	24.57	24.00	2.545	0.32	6.77	0.05	24.20	0.02	23.96	0.14	0.505	2.29	1
2	94	235	2.171	0.30	7.74	0.04	24.19	0.01	23.62	0.06	0.389	20.7	1
2	465	106	1	5.75	24.65	23.86	1.922	0.30	5.73	0.04	24.41	0.02	23.87	0.11	0.251	23.8	1
2	426	76	1.170	0.38	3.05	0.02	25.23	0.04	24.70	0.23	0.784	11.6	1
2	88	133	1.121	0.28	3.23	0.03	25.04	0.03	24.20	0.15	0.857	6.18	2
3	28	358	12	36.3	22.36	21.57	4.864	0.32	48.9	1.80	22.41	0.01	21.41	0.06	0.015	17.7	1
3	46	640	9	21.5	22.80	22.60	3.013	0.36	21.6	0.21	22.87	0.01	22.44	0.13	0.003	5.54	1
3	162	672	11	13.3	24.04	23.28	2.678	0.32	13.1	0.17	24.00	0.02	23.34	0.12	0.690	26.1	1
3	155	119	2	2.16	24.2	...	2.578	0.28	2.18	0.01	24.08	0.02	24.13	0.15	0.615	10.5	1
3	229	592	8	18.5	22.2	...	2.572	0.14	18.7	0.24	22.19	0.01	22.20	0.03	0.068	37.2	1
3	63	174	4	...	22.4	...	2.176	0.22	75.0	5.62	22.33	0.01	20.26	0.01	0.012	38.9	2
3	537	91	1	4.26	24.98	24.59	2.063	0.44	4.30	0.02	24.96	0.02	24.61	0.19	0.402	15.5	1
3	142	742	1.980	0.41	6.63	0.04	24.32	0.02	24.30	0.10	0.070	21.9	1
3	197	651	10	10.5	24.56	23.76	1.939	0.33	10.6	0.13	24.54	0.02	23.99	0.08	0.482	21.2	1
3	535	119	3	...	24.6	...	1.888	0.67	417.	...	24.36	0.27	21.79	0.03	0.541	303.	2
3	167	383	7	6.16	25.13	24.63	1.776	0.40	6.12	0.06	24.97	0.03	24.47	0.14	0.766	7.77	1
3	161	215	6	2.63	23.9	...	1.697	0.14	2.55	0.02	23.70	0.02	23.69	0.14	0.648	7.56	1
3	250	748	1.451	0.09	7.87	0.14	22.92	0.01	23.25	0.07	0.598	2.44	2
3	150	355	1.385	0.50	3.33	0.02	25.43	0.05	24.44	0.38	0.725	4.99	1
3	208	86	1.376	0.07	53.1	11.0	22.24	0.02	22.78	0.05	0.260	15.0	2
3	390	560	1.239	0.13	77.3	5.73	22.12	0.01	20.62	0.01	0.250	57.3	2
3	177	572	1.066	0.30	2.73	0.02	25.34	0.04	25.24	0.23	0.878	19.2	2
3	187	121	1.028	0.13	2.16	0.01	23.55	0.02	23.43	0.08	0.781	17.1	2
3	66	185	5	...	25.3	...	0.410	24.44	0.03	24.42	0.30
4	515	162	11	42.0	22.33	21.40	6.604	0.40	38.1	0.43	22.40	0.01	21.69	0.03	0.237	7.25	1
4	377	307	6	4.512	0.18	90.5	6.68	20.40	0.01	20.53	0.02	0.002	20.2	2
4	476	129	10	18.0	23.14	22.77	4.153	0.39	18.2	0.15	23.16	0.01	22.90	0.04	0.279	27.7	1
4	391	327	7	37.0	22.70	21.79	4.067	0.40	35.9	0.29	22.65	0.01	22.01	0.04	0.112	17.4	1
4	400	152	8	35.2	22.72	22.17	4.018	0.37	35.0	0.30	22.79	0.01	22.17	0.04	0.057	15.4	1
4	792	266	18	26.8	23.20	22.25	3.334	0.25	27.1	0.74	23.18	0.01	22.46	0.04	0.504	16.7	1
4	561	580	13	1.35	22.9	...	3.193	0.24	5.63	0.03	22.70	0.02	23.22	0.13	0.450	3.70	2
4	653	351	16	16.0	23.76	23.11	3.189	0.33	15.1	0.23	23.51	0.01	22.97	0.05	0.358	30.7	1
4	600	134	14	22.0	23.42	22.62	2.924	0.33	23.5	0.29	23.44	0.01	22.74	0.05	0.294	11.5	1
4	639	662	15	20.6	23.36	22.31	2.846	0.31	20.6	0.26	23.39	0.01	22.67	0.05	0.628	16.1	1
4	558	173	12	1.00	23.7	...	2.603	0.21	9.85	0.12	23.52	0.02	23.71	0.14	0.908	6.71	2
4	537	529	2.057	0.07	4.46	0.04	21.74	0.01	22.17	0.04	0.869	5.00	2
4	372	329	5	4.34	25.1	...	1.773	0.46	4.32	0.03	24.91	0.03	24.70	0.23	0.717	2.21	1
4	365	33	1.403	0.30	13.7	0.14	23.77	0.02	23.19	0.05	0.526	38.4	1
4	223	354	1.375	0.48	3.88	0.02	25.24	0.04	24.85	0.41	0.779	25.5	1
4	432	758	9	7.12	24.65	24.29	1.345	0.41	6.90	0.03	24.59	0.02	24.11	0.13	0.477	7.98	1
4	759	658	17	7.35	24.85	24.34	1.339	0.33	7.50	0.05	24.62	0.02	24.35	0.16	0.050	15.3	1
4	234	235	4	5.18	24.85	24.51	1.181	0.29	5.11	0.05	24.82	0.02	24.51	0.18	0.309	9.64	1
4	707	367	1.136	0.33	4.24	0.03	24.99	0.03	24.33	0.17	0.832	3.86	2
4	726	489	1.127	0.44	4.11	0.04	25.12	0.04	24.80	0.22	0.703	5.66	1
4	99	249	1.114	0.09	55.1	2.20	21.46	0.01	19.96	0.01	0.057	37.4	2
4	514	528	1.029	0.18	3.89	0.06	24.19	0.02	24.10	0.22	0.368	3.31	2
4	156	345	1.000	0.41	60.4	13.3	24.84	0.08	24.33	0.13	0.339	2.12	2
4	129	359	0.998	0.37	2.76	0.02	25.47	0.04	24.87	0.19	0.672	10.7	2
4	158	595	0.963	0.27	4.60	0.05	25.09	0.03	24.66	0.19	0.656	3.89	2
4	152	307	0.944	0.17	7.08	0.09	24.47	0.02	23.92	0.11	0.436	3.57	2
4	708	252	0.919	0.18	3.74	0.04	24.97	0.04	22.76	0.07	0.943	17.1	2

TABLE 1
(Continued)

Chip	x	y	ID	P	$\langle V \rangle$	$\langle I \rangle$	L	A	P	σ	$\langle V \rangle$	σ	$\langle I \rangle$	σ	$R_{2,1}$	P_2	Q
4	142	103	0.911	0.20	38.9	4.11	24.83	0.03	22.81	0.04	0.328	16.8	2
4	186	724	2	5.84	24.80	24.43	0.911	0.26	5.84	0.04	24.79	0.02	24.39	0.12	0.024	32.0	1
4	192	283	3	...	24.2	...	0.719	21.10	0.01	19.31	0.01
4	24	153	1	3.67	24.83	24.74	0.529	25.35	0.05	25.20	0.14

sents the previous authors' confidence in the detection, as I have not independently derived a period and light curve for this star. In the present analysis the star at the position of C4-V1 had an L value of 0.529. Among the 26,937 stars in my sample that appeared in at least 16 frame pairs, 705 had L values larger than 0.529, so in some sense this star could be regarded as a physical variable at the 97.4% confidence level. However, the increasingly high rate of false detections in comparison with the frequency of good-quality variables as L decreases suggests that in a Bayesian sense the statistical significance of the variation in C4-V1 is much lower than this. The relatively short period inferred for this star opens the possibility that it has simply been possible to find a permutation of the observed magnitudes which can produce the appearance of a reasonable light curve from random noise. All the remaining 36 variable candidates identified by Saha et al. had L indices from the present analysis equal to 0.911 or greater, based on the 43 images from 1992.

Table 1 summarizes the light-curve parameters for all high- and medium-confidence variable candidates based on the 43-frame analysis. Here, columns 1, 2, and 3 give, respectively, the chip and the (x, y) coordinates where the variable candidate may be found (in the coordinate system of the first exposure of the field, W0UA0101T, obtained 1992 January 27 02:47:17 UT); columns 4, 5, 6, and 7 list the identification number, period, and visual and infrared magnitudes assigned by Saha et al.; L is the variability index defined in the previous section; and A is the fitted fundamental-harmonic visual semi-amplitude. These are followed by the period and flux-weighted mean visual and infrared magnitudes derived from the present analysis, along with their formal standard errors. Finally, $R_{2,1}$ is the ratio of the score of the second-best light-curve fit to that of the best fit, as described above; P_2 is the value of the second-best period, and Q is a summary of the subjective light-curve and image-quality scores—the larger of the two is given in the table. Individual candidates are listed in order of decreasing L value, chip by chip.

Figure 3 shows a comparison of the periods found by Saha et al. with those determined here, for stars in common. The two wildly discrepant points at the left-hand side of the figure are their stars C4-V12 and C4-V13, which they identified as possible eclipsing binaries with periods of 1.00 and 1.35 days, respectively. The present reductions were not permitted to consider trial periods shorter than 2 days, so alternative—likely spurious—periods within the allowed range were settled on instead. Apart from these two, the only candidate variable for which I find an optimum period significantly different from that of Saha et al. is their star C3-V12, for which they obtain $P = 36.3$ days, while I estimate

$P = 48.9 \pm 1.8$ days. Figure 4 shows my reduced visual magnitudes for this star phased on both periods; the 49-day period seems at least as plausible as the 36-day period.

Figures 5 and 6 are, respectively, the V - and I -band period-magnitude relations for all candidate variables detected here. The symbol types are for the most part the same as in Fig. 2: filled circles are stars found by Saha et al. which were independently recovered here, while empty circles are candidate variables found here but not by Saha et al., and large symbols have subjective quality class 1 while smaller symbols have quality class 2. Here, however, '+'s refer to stars which were identified as candidate variables by Saha et al., but for which they were unable to determine a period, while the present analysis *did* find an acceptable period. In each diagram a line segment shows the effect of adopting Saha et al.'s period and magnitude for C3-V12, as opposed to those found here. The two small filled circles just to the left of the middle of the diagram are the two probable eclipsing binaries which may have true periods near .1 day. The two large filled circles near $\log(P)=0.4$, $V \approx 24$ are C3-V2 and C3-V6 of Saha et al., which they labeled "?" and "Eclipsing," respectively. Some of the other points which lie to the left of the main band of the P-L relation may also be eclipsing binaries, and one or two may even be Cepheids pulsating in an overtone, but note that all have subjective quality class 2. The three points which lie far below and to the right of the P-L relation are also all of quality class 2; they may be blunders, they may have incorrect periods, or they may be some other type of variable, such as W Virginis

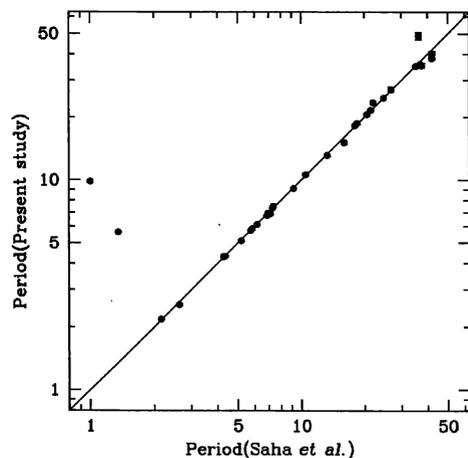


FIG. 3—A comparison of the periods estimated by the automatic algorithm of the present study, with those found by Saha et al. A line of unit slope and zero intercept is also indicated.

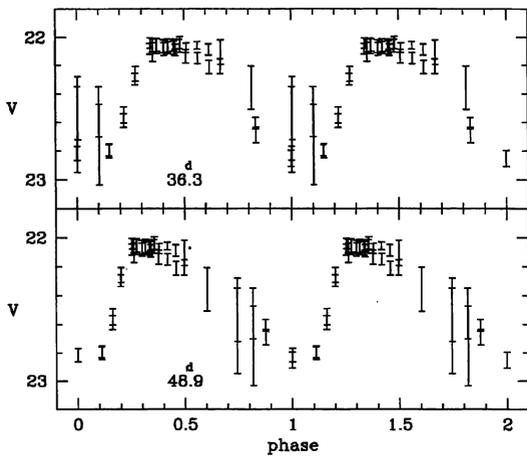


FIG. 4—Two alternative phasings of the light-curve data for the probable Cepheid C3-V12 (in the system of Saha et al. 1994): (upper panel) phased on the period of 36.3 days found by Saha et al.; (lower panel) phased on the period of 48.9 days found here.

stars, perhaps. Note that two of the stars for which Saha et al. were unable to assign periods are here placed plausibly close to the P-L relation, even though the estimated periods are considerably longer than the 46-day observing window—these periods were selected out by finding the best match possible between the available data string and a *part* of a template light curve. The large (quality class 1) filled circle which lies more than a magnitude above the P-L relation at

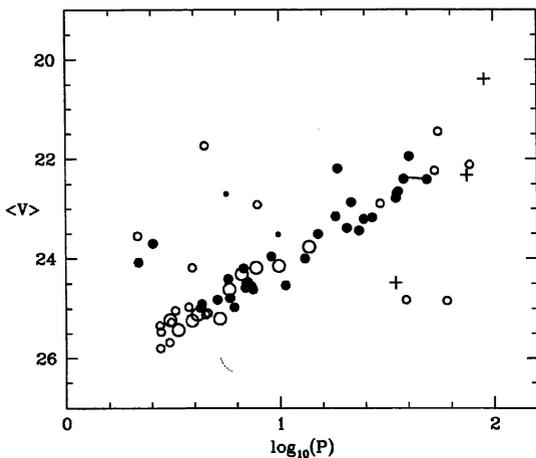


FIG. 5—V-band period-magnitude relation for variable candidates in IC 4182, based on the periods and flux-weighted mean magnitudes determined here. Filled circles represent candidate variables previously identified by Saha et al. (1994) while empty circles represent new candidates identified here. In each case, larger symbols represent variables with a high degree of subjective confidence, while smaller symbols represent candidates with a lower confidence level. Plus signs represent probable long-period variables which were identified by Saha et al., but for which they determined no periods; they are plotted with the periods and mean magnitudes estimated from the present analysis. A line connects the present determinations for the variable C3-V12 with those found by Saha et al.; the end of the line segment corresponding to the Saha et al. results lies at about the 10 o'clock position on the filled circle representing a different variable candidate.

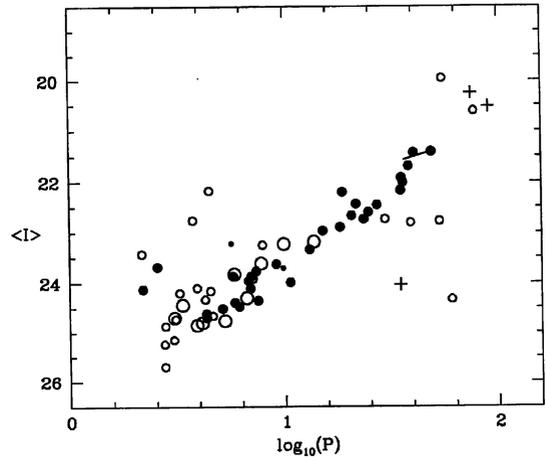


FIG. 6—Same as Fig. 5, but for the I-band results.

$\log(P)=1.27$ ($P=18.7\pm 0.24$ days) $\langle V \rangle = 22.19 \pm 0.01$ is C3-V8; the values obtained by Saha et al. for this star are virtually identical: $P=18.5$ days, $\langle V \rangle = 22.2$. It is therefore possible that this star truly is abnormally bright for its period. Note, however, that Saha et al. quoted only one decimal place for the mean visual magnitude of this star, designated it “Cepheid?” and noted “superposed stars?.” However, I gave it subjective quality classes “1” for both the appearance of the light curve and the appearance of the image.

Apart from the individual stars discussed in the previous paragraph, the remaining variable candidates form pleasantly tight period-magnitude relationships, including many of the quality class 2 variables found here for the first time. The diagrams argue plausibly that this technique has been effective at finding numerous true Cepheids in IC 4182 with periods as short as 2.5 days. Furthermore, the tightness of the sequence even at the shortest periods suggests that my aforementioned makeshift—employing equal-amplitude sinusoids

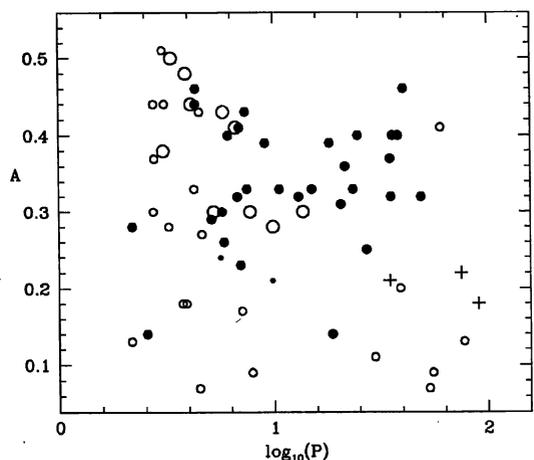


FIG. 7—Period-amplitude relation for the candidate variables in IC 4182, based upon results of the present analysis. Symbol types are as in Figs. 5 and 6.

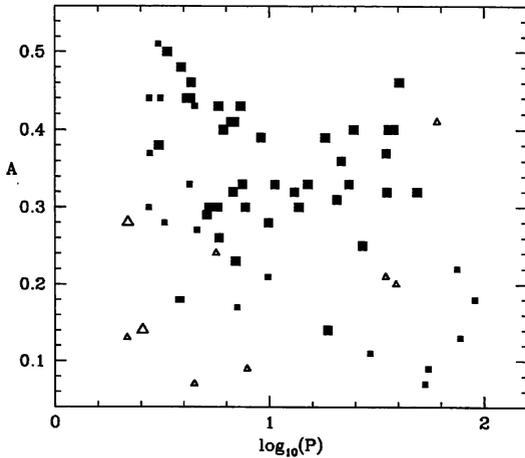


FIG. 8—Same as Fig. 7, except that here filled squares represent stars lying within ± 1.25 mag of the mean period–magnitude relation for IC 4182, whereas empty triangles represent stars lying outside that range. Again, larger symbols represent stars with higher subjective confidence levels, smaller symbols those with lower confidence levels.

for the V - and I -band light curves when the starting guess for the period was less than 7 days—has not seriously corrupted the derivation of the relevant light-curve parameters.

Figure 7 is a period–amplitude plot for all the variable candidates from the present study; symbol types are as in Figs. 5 and 6. Here we see that the new variable candidates which were detected here but *not* listed by Saha et al. (empty circles) tend to have either short periods (hence faint apparent magnitudes) or small amplitudes. The separation is not absolute, however: some of the new candidates have properties within the range where Saha et al. were successfully finding variables. Figure 8 repeats the information in Fig. 7, except that here the symbol types have been recoded: filled

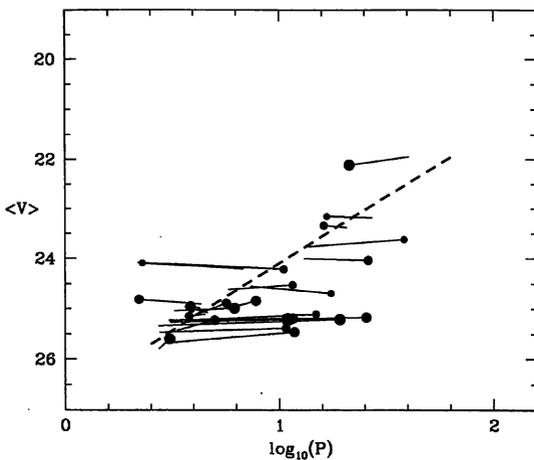


FIG. 9—The period–magnitude relation that would be obtained for IC 4182 if the second-best period and mean magnitude were adopted for each of those candidate variables which has a second solution with a score more than half as large as the first solution. Only candidates with $A > 0.22$ mag are included in this figure. Line segments connect each point with the position it would occupy if the results of its best solution were plotted instead.

squares represent stars that lie within ± 1.25 mag of the mean visual P-L relation shown in Fig. 5 (an interval chosen specifically to include the anomalously bright variable C3-V8), while open triangles represent those stars that lie outside this range. In fact, star C3-V8 is the large filled square at $\log(P)=1.27$, $A=0.14$ —the small perceived amplitude in this star provides strong, *objective* evidence that this star is, in fact, blended with one or more comparably bright companions, as originally suggested by Saha et al. I noted above that none of the sample of Milky Way and Magellanic Cloud Cepheids that were used to derive the template light curves had an amplitude $A < 0.194$; if we were to eliminate from the present sample all candidates with $A < 0.22$ we would reject C3-V8, plus six of the nine other outlying points (including one of the 1 day eclipsing variables) *without* rejecting any of the other quality class = 1 candidates. Among those stars that would be retained there are 20 candidates identified here but not by Saha et al., ten of them of quality class 1, and one long-period variable identified by Saha et al., for which the present analysis assigns a guesstimated period of 75 days. Only three outliers would be retained in the truncated sample: one of them the possible 1.35-day eclipsing binary, another a possible ~ 60 -day long-period variable or W Vir star (chip 4, $x=156$, $y=345$).

One disturbing feature of the period–amplitude diagram is the abundance of candidate variables in IC 4182 with periods less than 10 days and fundamental-harmonic visual semi-amplitudes $A > 0.4$ mag—a region of the plane completely empty in the calibrating sample of Galactic and Magellanic Cloud Cepheids (Appendix, Fig. 22 below). If upon further investigation this anomaly should persist, it might suggest some fundamental difference between the Cepheid populations of IC 4182 and the calibrating galaxies. In the worst of all possible cases, this might call into question the universality of the Cepheid period–luminosity relation.

Figure 9 illustrates how the visual P-L relation would appear if the period and mean magnitude of the *second*-best fit were used for each variable which had $R_{2,1} \geq 0.5$. In this plot, points represent only those stars which had $A > 0.22$, and the size of the circle is proportional to $R_{2,1}$. A line segment connects each point to the period and mean magnitude derived from the *best* fit. It is seen that the almost universal tendency is for the second-best solution to lie farther from the mean P-L relation than the best, even when $R_{2,1}$ is close to unity (the largest dots). The one notable exception is the star C3-V2, labeled “?” by Saha et al., which is most probably an eclipsing binary or first overtone Cepheid with a period of 2.2 days (according to both this study and Saha et al.) or, barely possibly ($R_{2,1}=0.615$), it is a fundamental-mode Cepheid with a period of 10.5 days if the second-best solution is adopted—but probably not. I conclude that the fit-scoring scheme outlined above is quite effective at choosing the best period from the list of available possibilities for each star.

The photometry of all detected stars, and the template-fitting analysis of all detected variable candidates, was repeated on the basis of all 47 available observations of the galaxy, including the cosmic-ray split V and I pairs from 1993 January 24. The results of this analysis are reported in

TABLE 2
Cepheid Parameters Derived from 47 Observations in 1992-1993

Chip	x	y	ID	P	(V)	(I)	L	A	P	σ	(V)	σ	(I)	σ	R _{2,1}	P ₂	Q
1	49	62	6	42.0	21.87	21.19	7.855	0.46	40.5	0.90	21.95	0.01	21.43	0.02	0.855	21.3	1
1	314	78	4	24.7	23.29	22.45	4.310	0.39	25.2	0.06	23.22	0.01	22.60	0.04	0.829	27.1	1
1	330	67	5	9.20	24.10	23.55	2.614	0.39	9.19	0.01	23.97	0.02	23.61	0.06	0.560	10.4	1
1	576	454	2.201	0.28	10.1	0.01	24.15	0.01	23.26	0.04	0.472	18.6	1
1	572	181	1.876	0.42	5.82	0.00	24.61	0.03	23.84	0.08	0.627	11.7	1
1	358	493	2	7.30	24.54	23.75	1.701	0.43	7.31	0.01	24.56	0.02	23.74	0.08	0.724	7.82	1
1	336	377	3	1.445	0.21	35.0	0.44	24.49	0.02	24.10	0.13	0.978	29.8	2
1	249	517	1	6.95	24.47	23.71	1.329	0.24	7.05	0.06	24.47	0.02	23.86	0.08	0.530	44.7	1
1	337	504	1.317	0.30	5.18	0.01	25.22	0.03	24.82	0.18	0.522	14.7	1
1	355	188	1.290	0.40	4.42	0.00	25.13	0.03	24.22	0.14	0.566	9.86	2
1	393	297	1.236	0.44	3.10	0.00	25.26	0.04	24.55	0.23	0.897	9.66	2
1	334	658	1.180	0.41	2.70	0.00	25.79	0.04	25.75	0.68	0.759	7.92	2
1	350	578	1.100	0.50	3.01	0.01	25.66	0.05	25.07	0.29	0.788	11.6	2
1	592	260	1.088	0.16	85.7	1.06	22.97	0.04	22.74	0.05	0.954	49.0	2
2	412	190	2	37.5	22.86	21.85	6.994	0.32	34.1	0.04	22.69	0.01	21.91	0.03	0.980	37.8	1
2	94	235	1.961	0.29	7.76	0.04	24.19	0.01	23.63	0.06	0.182	18.6	1
2	245	109	3	6.88	24.57	24.00	1.876	0.33	7.00	0.01	24.20	0.02	23.81	0.13	0.581	14.6	1
2	465	106	1	5.75	24.65	23.86	1.846	0.30	5.73	0.04	24.41	0.02	23.87	0.10	0.304	21.9	1
2	426	76	1.109	0.38	3.05	0.01	25.23	0.04	24.74	0.20	0.776	11.5	1
3	28	358	12	36.3	22.36	21.57	5.032	0.32	50.1	0.11	22.41	0.01	21.56	0.05	0.955	58.7	1
3	162	672	11	13.3	24.04	23.28	2.686	0.32	12.9	0.06	24.02	0.02	23.42	0.10	0.930	14.4	1
3	46	640	9	21.5	22.80	22.60	2.654	0.37	21.4	0.18	22.86	0.01	22.47	0.05	1
3	229	592	8	18.5	22.2	...	2.456	0.15	18.9	0.05	22.19	0.01	22.17	0.03	0.651	17.1	1
3	63	174	4	...	22.4	...	2.436	0.35	102.	0.42	22.48	0.02	20.29	0.01	0.716	78.8	2
3	142	742	2.144	0.41	6.64	0.02	24.32	0.02	24.27	0.08	0.040	19.9	1
3	537	91	1	4.26	24.98	24.59	1.862	0.44	4.30	0.02	24.96	0.03	24.61	0.18	1
3	155	119	2	2.16	24.2	...	1.853	0.28	2.18	0.01	24.08	0.02	24.13	0.14	0.609	10.5	1
3	197	651	10	10.5	24.56	23.76	1.727	0.33	10.7	0.03	24.54	0.02	24.00	0.07	0.741	9.82	1
3	535	119	3	...	24.6	...	1.721	0.67	413.	...	24.36	0.27	21.79	0.03	0.034	6.60	2
3	161	215	6	2.63	23.9	...	1.366	0.16	7.80	0.01	23.74	0.02	23.78	0.11	0.373	2.59	1
3	167	383	7	6.16	25.13	24.63	1.360	0.41	6.30	0.01	24.97	0.03	24.52	0.11	0.501	12.7	1
3	250	748	1.357	0.08	8.10	0.01	22.92	0.01	23.19	0.05	0.453	7.43	2
3	390	560	1.352	0.15	92.1	5.72	22.19	0.02	20.65	0.01	0.527	75.3	2
3	152	242	1.324	0.04	11.8	0.22	20.53	0.01	20.80	0.02	0.509	16.3	2
3	150	355	1.312	0.50	3.31	0.00	25.43	0.05	24.80	0.27	0.737	2.34	1
3	208	86	1.206	0.07	53.1	11.0	22.24	0.02	22.78	0.05	0.264	15.0	2
3	177	572	1.117	0.30	2.73	0.01	25.32	0.04	24.87	0.18	2
3	187	121	1.032	0.13	2.16	0.01	23.54	0.02	23.43	0.08	0.778	17.1	2
3	659	308	0.946	0.16	2.58	0.02	25.24	0.03	23.73	0.07	0.602	3.18	2
3	171	483	0.910	0.14	92.6	9.11	22.46	0.02	20.94	0.01	0.182	73.5	2
3	66	185	5	...	25.3	...	0.394	24.46	0.03	24.57	0.16
4	515	162	11	42.0	22.33	21.40	6.002	0.40	42.0	0.44	22.32	0.01	21.64	0.03	0.810	48.0	1
4	391	327	7	37.0	22.70	21.79	4.507	0.40	35.8	0.04	22.64	0.01	21.97	0.03	0.848	39.9	1
4	400	152	8	35.2	22.72	22.17	4.141	0.38	35.6	0.29	22.79	0.01	22.20	0.04	0.656	59.9	1
4	377	307	6	4.141	0.20	105.	2.22	20.40	0.01	20.52	0.02	0.563	82.6	2
4	476	129	10	18.0	23.14	22.77	3.723	0.39	18.2	0.15	23.16	0.01	22.90	0.04	0.277	27.7	1
4	639	662	15	20.6	23.36	22.31	3.055	0.32	20.4	0.12	23.38	0.01	22.64	0.04	0.942	21.6	1
4	792	266	18	26.8	23.20	22.25	3.052	0.25	26.9	0.17	23.19	0.01	22.45	0.04	0.976	28.6	1
4	561	580	13	1.35	22.9	...	2.951	0.10	4.04	0.01	22.58	0.02	22.90	0.08	0.914	2.63	2
4	653	351	16	16.0	23.76	23.11	2.948	0.33	15.2	0.04	23.51	0.01	22.98	0.04	0.698	14.1	1
4	600	134	14	22.0	23.42	22.62	2.584	0.33	23.5	0.29	23.44	0.01	22.74	0.04	1
4	558	173	12	1.00	23.7	...	2.257	0.23	6.66	0.04	23.48	0.02	23.60	0.15	0.949	3.26	2
4	537	529	1.847	0.07	4.50	0.01	21.74	0.01	22.18	0.04	0.544	3.65	2
4	372	329	5	4.34	25.1	...	1.739	0.46	4.36	0.00	24.92	0.03	24.66	0.14	0.564	8.80	1
4	432	758	9	7.12	24.65	24.29	1.607	0.40	7.04	0.04	24.57	0.02	24.00	0.11	0.372	14.8	1
4	223	354	1.484	0.48	3.82	0.00	25.22	0.03	23.82	0.21	0.652	19.6	1
4	759	658	17	7.35	24.85	24.34	1.416	0.33	7.48	0.01	24.62	0.02	24.30	0.11	0.061	14.6	1
4	156	345	1.315	0.43	62.4	0.67	24.85	0.04	24.24	0.12	0.961	76.9	2
4	708	252	1.304	0.41	87.0	4.18	24.65	0.03	22.57	0.04	0.865	69.5	2
4	152	307	1.266	0.16	6.81	0.11	24.47	0.02	24.00	0.11	0.595	3.58	2
4	365	33	1.209	0.30	13.7	0.14	23.77	0.02	23.19	0.05	0.528	38.4	1
4	726	489	1.154	0.45	4.01	0.00	25.08	0.04	24.73	0.19	0.621	11.7	1
4	707	367	1.100	0.32	4.26	0.01	24.99	0.03	24.27	0.13	0.850	3.84	2
4	514	528	1.032	0.20	3.95	0.00	24.20	0.02	24.13	0.16	0.048	8.17	2
4	99	249	1.027	0.15	82.7	6.76	21.53	0.02	19.98	0.01	0.164	324.	2

TABLE 2
(Continued)

Chip	x	y	ID	P	$\langle V \rangle$	$\langle I \rangle$	L	A	P	σ	$\langle V \rangle$	σ	$\langle I \rangle$	σ	$R_{2,1}$	P_2	Q
4	234	235	4	5.18	24.85	24.51	0.984	0.29	5.10	0.02	24.82	0.02	24.52	0.16	0.391	10.2	1
4	525	337	0.975	0.17	117.	14.8	23.87	0.04	22.24	0.04	0.737	4.77	2
4	190	270	0.948	0.41	4.03	0.02	25.13	0.05	23.48	0.12	0.897	6.98	2
4	731	448	0.929	0.40	4.34	0.01	24.97	0.04	24.66	0.19	0.943	3.53	2
4	129	359	0.921	0.37	2.76	0.00	25.47	0.04	24.76	0.18	0.292	6.18	2
4	311	212	0.916	0.23	5.71	0.04	24.68	0.02	24.36	0.11	0.084	9.23	2
4	142	103	0.915	0.20	38.9	4.09	24.83	0.03	22.81	0.04	0.336	16.8	2
4	186	724	2	5.84	24.80	24.43	0.822	24.74	0.03	24.30	0.12
4	192	283	3	...	24.2	...	0.710	21.10	0.01	19.31	0.01
4	24	153	1	3.67	24.83	24.74	0.523	25.35	0.05	25.20	0.14

Table 2, where the columns are the same as in Table 1. I found the conclusions scarcely changed: the L scores of C3-V5, C4-V1, and C4-V3 were 0.394, 0.523, and 0.710, respectively, so their apparent variability would still appear to be statistically insignificant according to the present analysis, as they were when it was based on 43 observations. C4-V2 (identified as ‘‘Cepheid’’ by Saha et al., with a period of 5.84 days; considered quality class 1 here) has also dropped just below the nominal variability limit of 0.900, with L decreasing from 0.911 to 0.822; all other Saha variables had L in excess of 0.980. Apart from these stars, two quality class 2 variables dropped from the sample: (chip, x , y) = (2, 88, 133) and (4, 158, 595), while five new quality class 2 candidates appeared: (3, 659, 308), (3, 171, 483), (4, 525, 337), (4, 190, 270), and (4, 731, 448). It seemed that in general the L values decreased overall when the four frames from 1993 were added to the 1992 data. I also noticed an increased star-to-star scatter among the nonvariable stars in the later frames as compared to the earlier. Presumably the increased random errors result from the fact that the 1993 pointing was translated by some 146 pixels (≈ 15 arcsec) and rotated by 3.7 deg relative to the 1992 observations, coupled with a less-than-perfect modeling of the spatial variation of the PSF. The larger standard errors of the 1993 magnitudes reduced the formal significance level of most detected variables.

The addition of the 1993 epoch yields a revised period of 50.1 ± 0.1 days for variable candidate (3, 28, 358) = C3-V12, which supports the period of 48.9 ± 1.8 days found above, rather than the 36-day period reported by Saha et al. The differences between the period found here based on 47 observations and those based on 43 observations for all candidates are summarized in Fig. 10. It is seen that the vast majority of periods have not changed significantly. It is also seen that the present algorithm tends systematically to underestimate the true periods of variables with periods longer than the time span of the observations. This is reasonable: with the string-length test not permitted to attempt trial periods longer than the time baseline of the observations, the iterated template fits must necessarily approach the variable’s true period from below. The convergence criterion is therefore more likely to be satisfied before the true period is reached than after. However, the case of candidate (4, 156, 345) for which a period of 60.4 ± 13.3 days was estimated from the 46.2-day baseline in 1992, and for which the de-

rived period became 62.4 ± 0.7 days upon addition of the 1993 epoch, demonstrates that a modest degree of extrapolation may sometimes be attempted. The two stars on the upper part of the unity line which have equal vertical and horizontal error bars are (3, 208, 86) and (4, 142, 103); for these two, and for a few other stars farther down the line, the rotation and translation between the 1992 and 1993 epochs shifted them off their chips, so no 1993 data were available. (I made no attempt to determine whether these stars had been shifted onto some other chip, so that I would not have to address the question of possible differences in the chip-to-chip zero points, or the possibility of different systematic errors in the various point-spread functions. Note that the L indices and the σ ’s for these stars may differ slightly in the two analyses, because the estimate of the standard errors—and hence, the relative weights—of the instrumental magnitudes have been influenced by the scatter found for *constant* stars, which was slightly different in the two analyses.)

Light curves for the five stars in Fig. 10 which lie off the

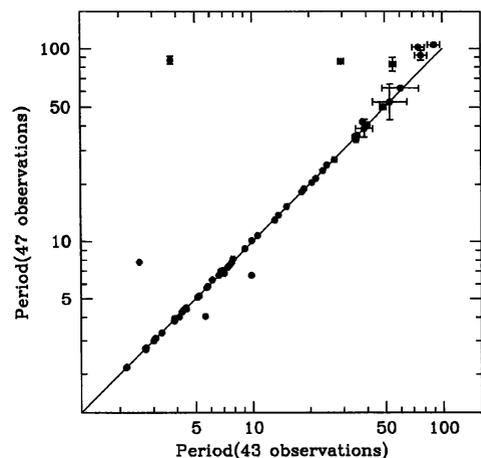


FIG. 10—A comparison of the periods found when all 47 observational epochs were included in the solution, with the results obtained when only the 43 epochs from 1992 are used. Error bars are plotted with every point, but in most cases they are lost within the symbol. A line with zero intercept and unit slope is plotted as well.

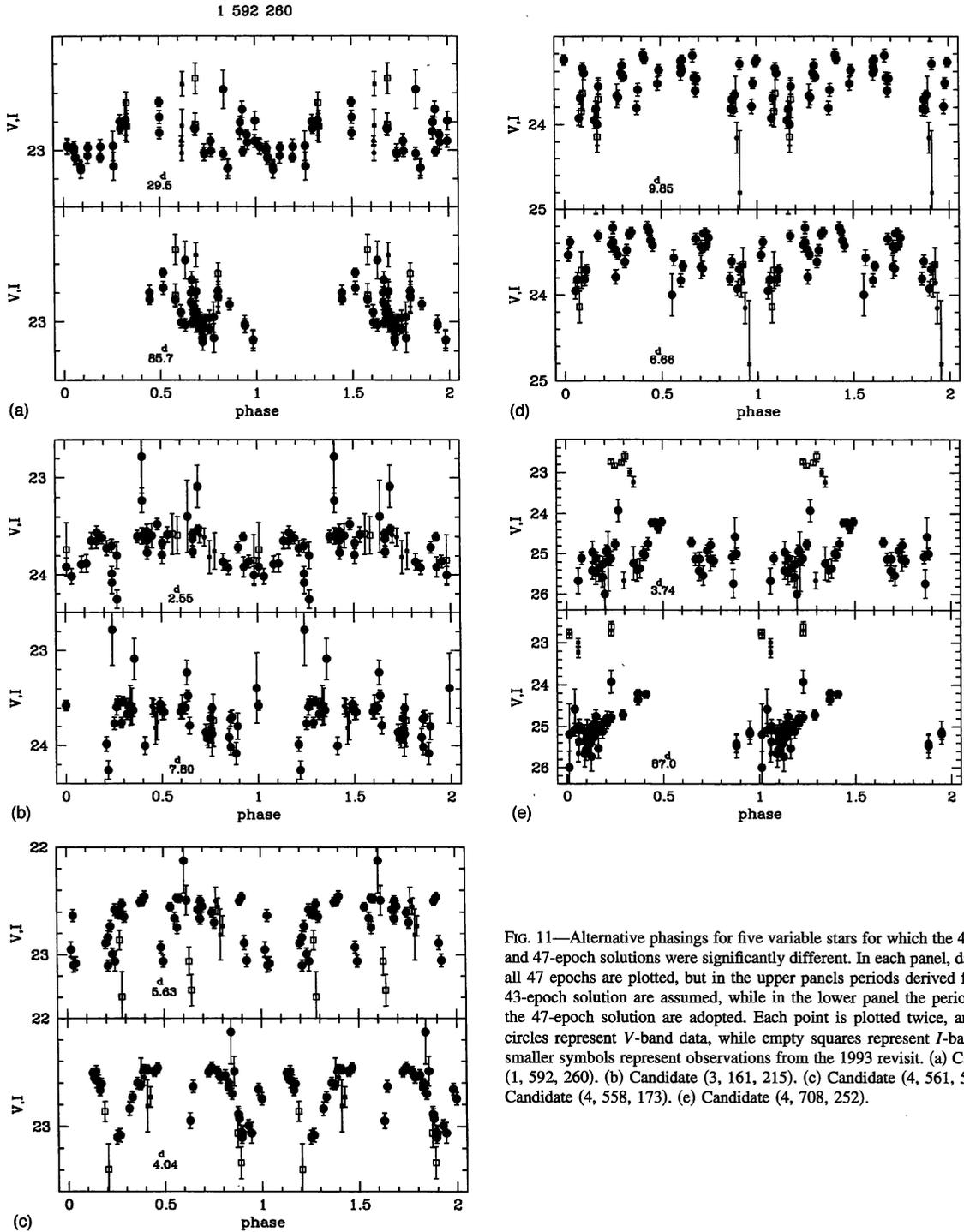


FIG. 11—Alternative phasings for five variable stars for which the 43-epoch and 47-epoch solutions were significantly different. In each panel, data from all 47 epochs are plotted, but in the upper panels periods derived from the 43-epoch solution are assumed, while in the lower panel the periods from the 47-epoch solution are adopted. Each point is plotted twice, and filled circles represent V-band data, while empty squares represent I-band data; smaller symbols represent observations from the 1993 revisit. (a) Candidate (1, 592, 260). (b) Candidate (3, 161, 215). (c) Candidate (4, 561, 580). (d) Candidate (4, 558, 173). (e) Candidate (4, 708, 252).

unity line and have $P(43 \text{ observations}) < 50$ days, are illustrated in Figs. 11(a)–11(e). In each case, the data from all available frames are plotted, but they are phased according to (upper) the 43-observation light-curve parameters, and (lower) the 47-observation light-curve parameters; the adopted period is noted near phase 0.5 in each panel. In each case, filled circles represent V-band data, and empty squares represent I-band data, and magnitudes from the 1993 revisit

are identified by smaller symbols. In all cases these light curves were assigned quality class 2, and it seems to me that there is little evidence that any of the periods has been convincingly demonstrated.

Finally, Fig. 12 illustrates the derived light curves for all variable candidates assigned quality class 1 in the 47-observation analysis which were not listed by Saha et al. In each case the upper panel portrays the data as a linear se-

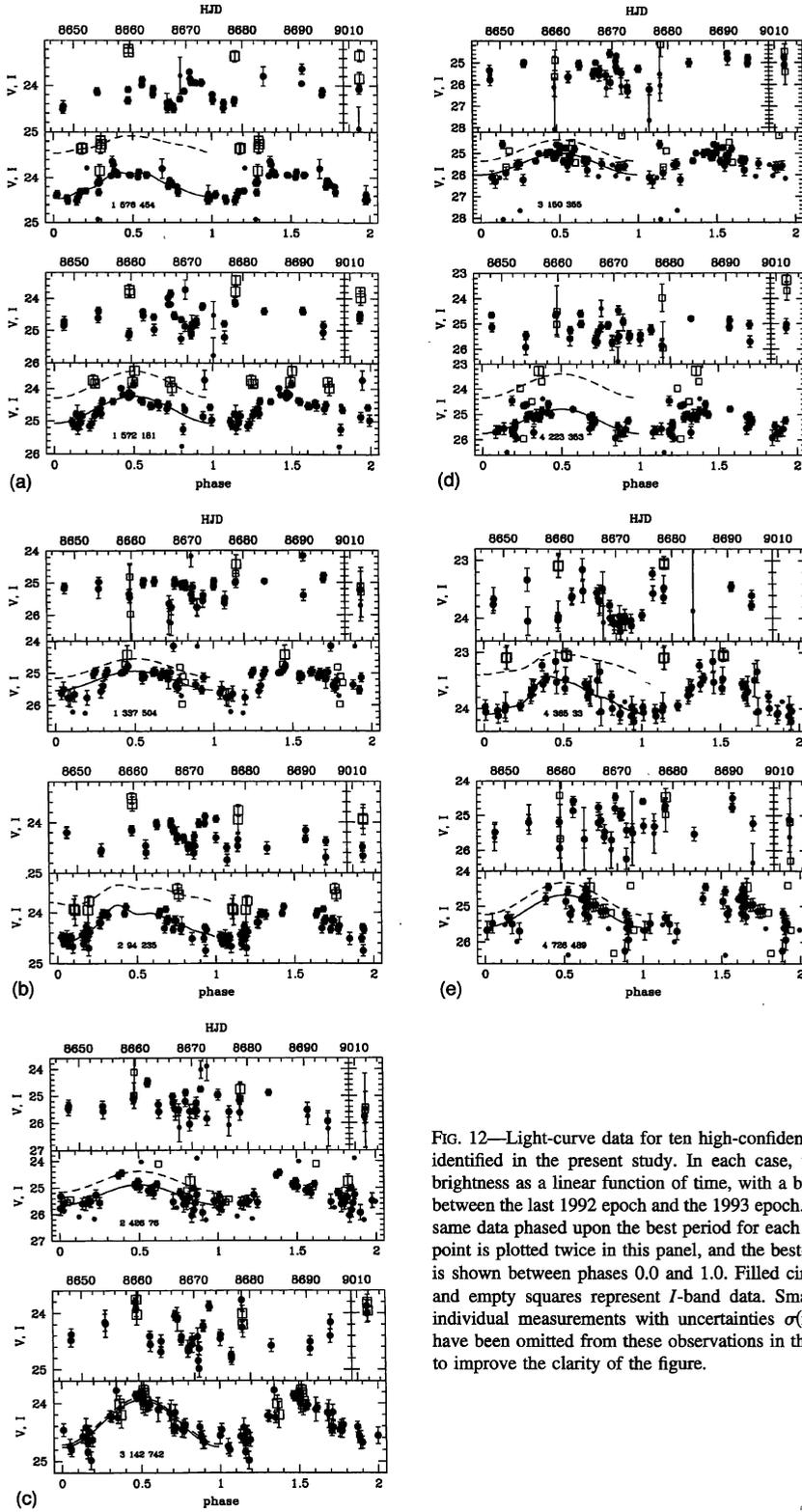


FIG. 12—Light-curve data for ten high-confidence new variable candidates identified in the present study. In each case, the upper panel represents brightness as a linear function of time, with a break in the horizontal scale between the last 1992 epoch and the 1993 epoch. The lower panel shows the same data phased upon the best period for each star given in Table 2; each point is plotted twice in this panel, and the best-fitting template light curve is shown between phases 0.0 and 1.0. Filled circles represent *V*-band data and empty squares represent *I*-band data. Smaller symbols are used for individual measurements with uncertainties $\sigma(\text{mag}) > 0.3$ mag; error bars have been omitted from these observations in the lower panel of each plot, to improve the clarity of the figure.

quence in time, with a break in the horizontal scale between the 1992 and 1993 observing seasons, while the lower panel shows the data phased according to the best period listed in Table 2. Again, filled circles have been used to represent

V magnitudes, and empty squares represent the *I* magnitudes. Low-weight observations, defined as those with $\sigma(\text{magnitude}) > 0.30$ mag are represented by smaller symbols, and in each lower panel the error bars have been left

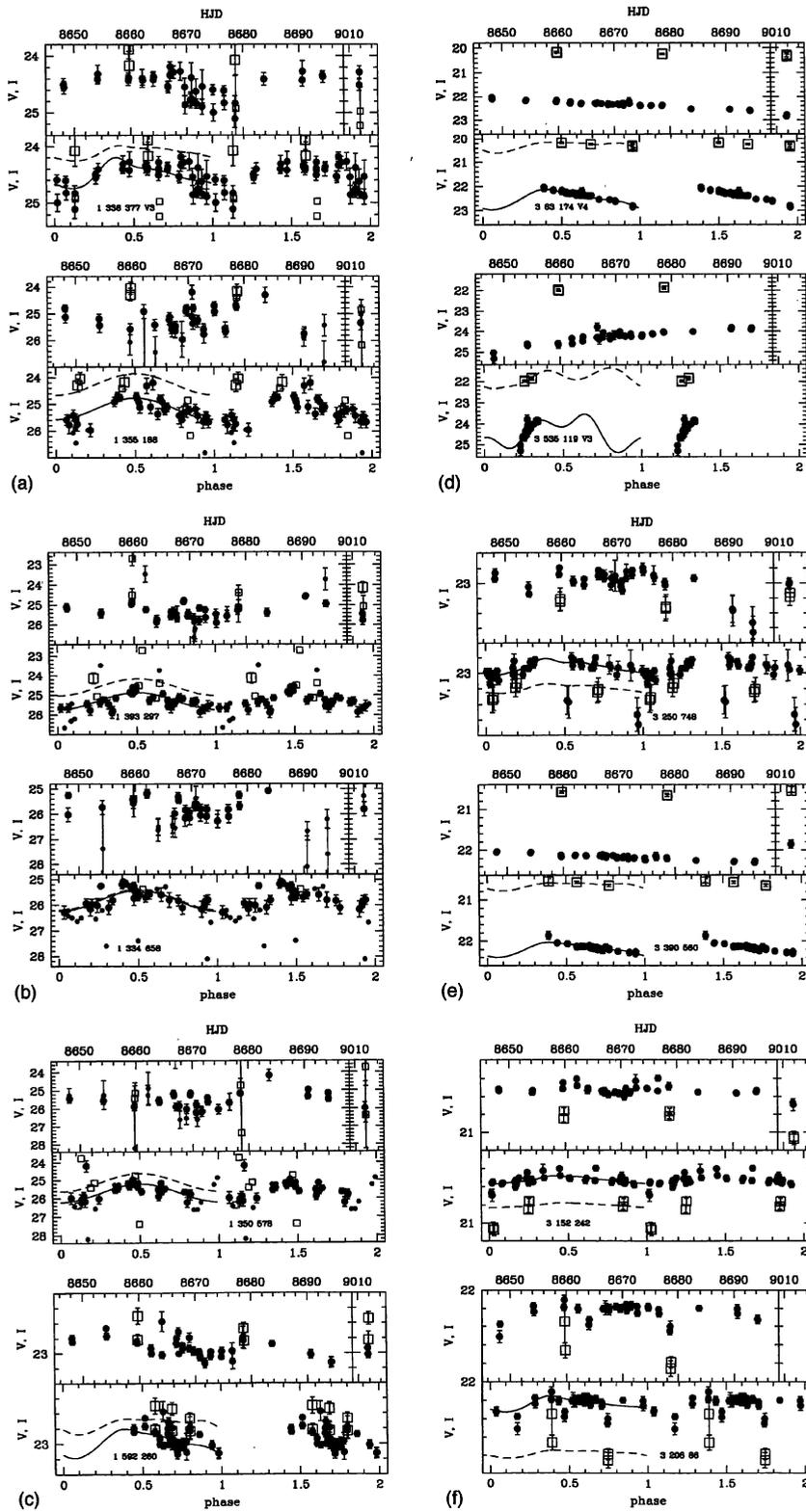


FIG. 13—Same as Fig. 12, except for 30 candidate variables with somewhat lower confidence levels, which either were identified here for the first time, or were identified by Saha et al. (1994), but for which they derived no period. In each case, on the vertical scale major ticks are at 1 mag intervals, and minor ticks are at 0.2 mag intervals.

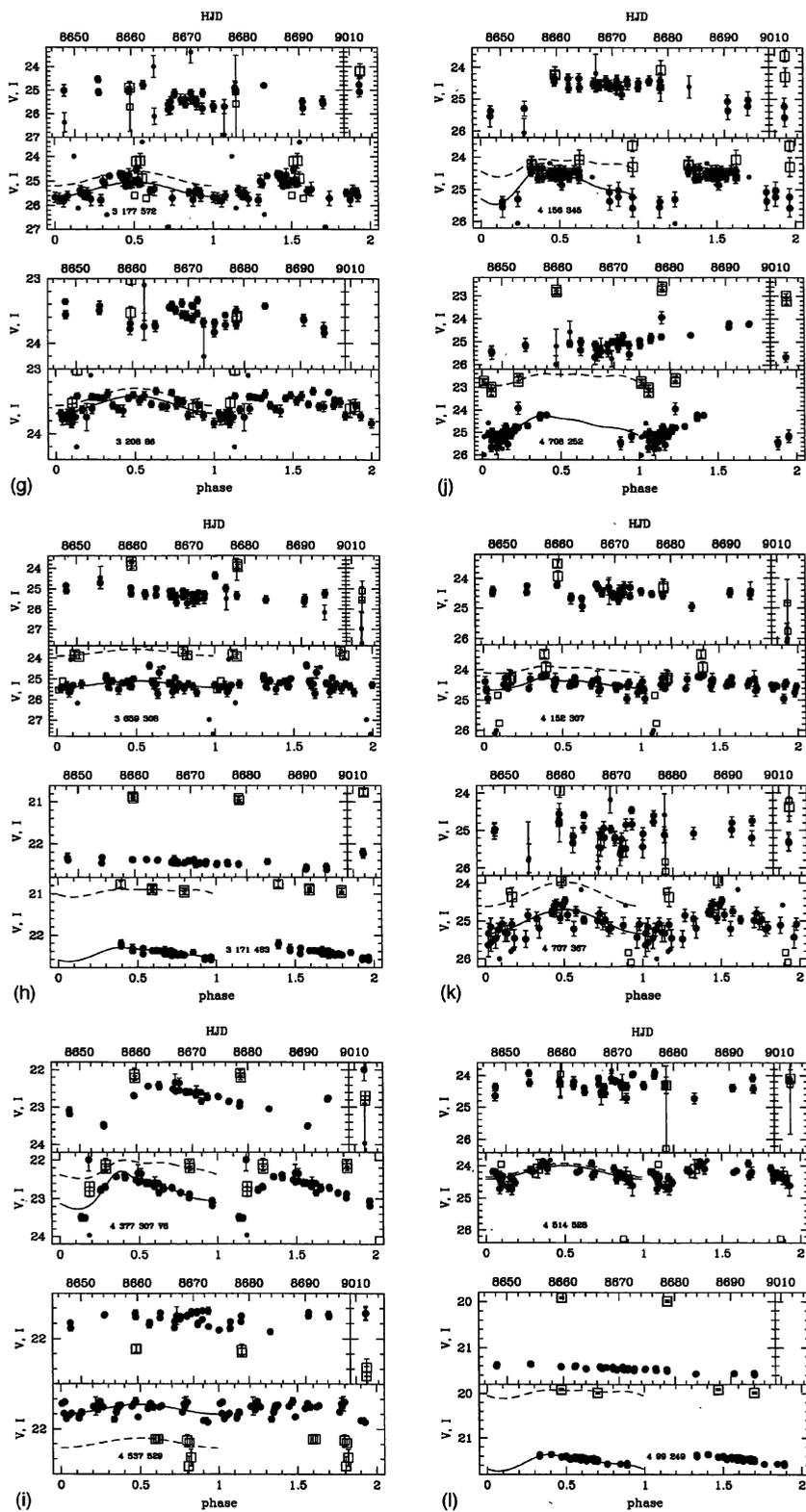


FIG. 13—(Continued)

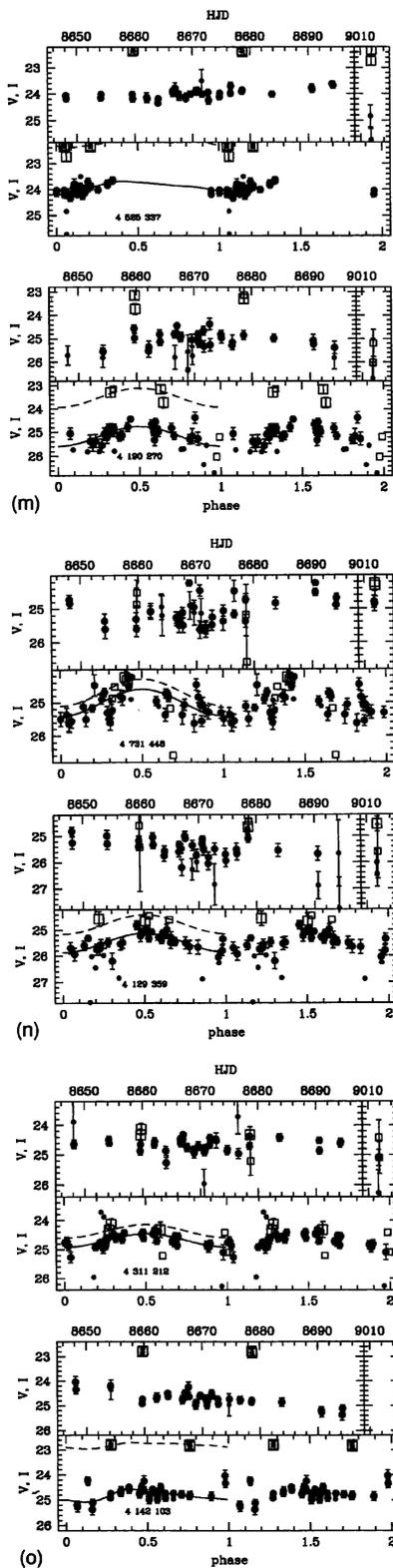


FIG. 13—(Continued)

off these points for better clarity. In the lower panels each point has been plotted twice, and the best-fitting template light curve is reproduced between phase = 0.0 and 1.0.

There is a lot of information in Fig. 12. For instance, note that candidate (3, 142, 742), with a period of 6.6 days and a sinusoidal light curve, has a $(V-I)$ color near 0.0, and is probably too blue to be a Cepheid. Conversely, candidate (2, 426, 76), with a period of 3.05 days and also fitted to a sinusoidal light curve, is somewhat redder, and furthermore the actual data for the star exhibit a steeper rising branch and longer, straighter descending branch than a true sinusoid. Thus, it is possible that this is an actual 3-day Cepheid. Candidate (2, 94, 235) shows how the inclusion of I -band data in both the candidate identification and the fitting of the template light curve can increase our confidence in the results: V - and I -band light curves corresponding to a single period and fixed template relations satisfy data in both bandpasses at once. This lends credibility to my claim that we can measure a precise flux-weighted $\langle I \rangle$ magnitude, even though effectively only two phases of the cycle have been sampled.

Figure 13 shows the same plots for all candidates which were assigned quality class 2 in the present 47-observation analysis, and were not listed by Saha et al., or for which they gave no period.

5. SUMMARY

This paper has presented a set of algorithms which is designed to identify high-quality variable-star candidates, and determine their light-curve parameters in an effective, robust, objective, and repeatable fashion. The method is intended to work as well as can be practically expected on small datasets with a moderately high level of contamination.

A comparison of the results of this algorithm with previous work on the same data by a large team of highly experienced astronomers suggests that these goals have been met. All of the previous researchers' high-quality candidates have been recovered and, with one exception, the same periods have been found by a completely impersonal computer algorithm—the sole interference on the part of an astronomer has been to impose the condition that periods shorter than 2 days are to be selected against, and amplitudes like those of Cepheids are to be preferred to those either much greater or much smaller. In the case of the one exception, where a significantly different period has been found here than was reported before, evidence has been presented that the algorithm's derived period is not worse, and may in fact be more correct than the previous one. The computer program has rejected four of the previous low-quality candidates (in the case of the 47-observation analysis; only three were rejected when exactly the same 1992 dataset as was used by Saha et al. 1994 was subjected to the algorithm), and has presented quantitative numerical indices which suggest that the candidates rejected are not indeed variable at a statistically significant confidence level. At the same time, the computerized algorithm has identified approximately ten new high-quality candidates and a somewhat larger number of new low-quality candidates, with statistical significance levels and quality indices within the same range of values as the previously known candidates. In a few cases, the algorithm has

made not unreasonable estimates of the periods and magnitudes of a few stars for which the previous investigators attempted no such determinations. It has been shown that objectively determined indices, such as the apparent amplitude of variation, can be used to separate high- from low-quality candidates. One can hope that, with further experimentation, such indices may eventually replace—or, at least, supplement in a major way—subjective quality classes assigned by the researcher. Finally, it has been demonstrated that when a good-quality light curve is available for one bandpass, a reliable flux-weighted average magnitude can usually be obtained for a second bandpass even when the phase coverage is poor.

I must stress, however, the reader should note that I am *not* justified in claiming that this computerized algorithm is *more* effective than a team of highly qualified astronomers. It was the purpose of Saha et al. to derive a Cepheid-based distance to the galaxy IC 4182. To that end, they did not need a *complete* sample of Cepheids; they needed a *clean* sample. This was achieved by rejecting from their sample and their publication any variable-star candidate in which they did not have absolute confidence. Their decisions were based partly on subjective and partly on quantitative criteria that were almost certainly quite different in detail from those imposed here. We have no way of knowing which or how many of the current “new” candidates were considered by the previous researchers, and then discarded for valid reasons that I have not considered. The most I can claim is that the present method has been shown to be competitive with a large team of experienced astronomers.

That said, I believe that certain advantages may be claimed for a computerized technique like the present one on largely philosophical grounds. (1) It is objective. Every operation, every decision has been explicitly defined. Once the programmer’s personal expertise and prejudice have been codified as mathematical equations and logical operations, the method works with absolute impartiality. (2) It is repeatable. Given the same data, the same answers will be obtained by an expert and a novice. Again given the same data, the same answer will be obtained this year as next, which is not always the case with a human being. (3) It is quantitative. It takes account of the numerical standard error of each measurement, weighting individual observations and defining formal confidence intervals for results with a level of fairness hard for a human being to match. (4) It is fast. Extensive tests employing large numbers of images to which synthetic Cepheids of known properties have been added can be reasonably undertaken to map out the detailed properties of the algorithm; conclusions based on these tests can be used to understand quantitatively the incompleteness of the real sample. When new data arrive, preliminary estimates of candidate lists and light-curve parameters can be quickly updated, using exactly the same methods and rules as before. (5) It is extensible. Alternate assumptions, alternate thresholds, alternate decision paths may be defined and applied to the same data; the resulting samples can be rigorously inter-compared. When a bug is found, it can be corrected, and precise quantitative implications for the results of previous analyses can be inferred. When our understanding of the de-

TABLE 3
Fourier Coefficients of Cepheid Light Curves

α_2	=	-0.210 ± 0.010	-0.091 ± 0.063	X	-0.333 ± 0.086	X^2
α_3	=	-0.102 ± 0.011	-0.218 ± 0.064	X	+0.058 ± 0.089	X^2
α_4	=	0.086 ± 0.010	-0.034 ± 0.026	X		
α_5	=	-0.071 ± 0.010	+0.028 ± 0.026	X		
β_2	=	1.8429 ± 0.105				if $X < -0.06$
	=	1.1796 ± 0.056	+0.506 ± 0.143	X		if $X \geq -0.06$
β_3	=	-0.030 ± 0.210				if $X < -0.06$
	=	-2.147 ± 0.116	+3.2175 ± 0.293	X		if $X \geq 0.06$
β_4	=	0.047 ± 0.076				
β_5	=	1.164 ± 0.093				
γ_1	=	0.608 ± 0.011	-0.022 ± 0.032	X		
γ_2	=	0.109 ± 0.011	+0.106 ± 0.032	X		
γ_3	=	-0.078 ± 0.011	-0.058 ± 0.032	X		
γ_4	=	0.046 ± 0.008				
δ_1	=	-0.190 ± 0.017	-0.254 ± 0.052	X		
δ_2	=	-1.340 ± 0.203				if $X < -0.06$
	=	-1.866 ± 0.128	+0.127 ± 0.349	X		if $X \geq -0.06$
δ_3	=	0.007 ± 0.277				if $X < -0.06$
	=	-2.946 ± 0.178	+4.823 ± 0.491	X		if $X \geq -0.06$
δ_4	=	-0.413 ± 0.168				

fining properties of Cepheids changes, the templates and selection criteria can be adjusted accordingly, and the analysis easily reperformed. One can envision the method gradually improving in proportion to the accumulated experience of many researchers.

I am grateful to Doug Welch, Nancy Silbermann, Barry Madore, Wendy Freedman, Horace Smith, Don Fernie, and the anonymous referee for their contributions to this work. In particular, this paper owes much to the excellent World-Wide Web sites managed by Doug Welch (<http://www.physics.mcmaster.ca/Cepheid/HomePage.html>) and Don Fernie (<http://ddo.astro.utoronto.ca/cepheids.html>).

APPENDIX: DEFINING THE TEMPLATE LIGHT CURVES

Since Cepheid light curves are periodic and extremely repeatable (except in the case of multiple-mode variables, which will be avoided here) it is convenient to describe them

TABLE 4
Cepheid Parameters Obtained during Template Derivation

Name	$\langle V \rangle$	σ	$\langle I \rangle$	σ	P	σ	A	σ_V	σ_I	n_V	n_I
Milky Way											
U Aql	6.451	0.006	5.238	0.024	7.0243	0.0007	0.319	0.040	0.019	39	39
SZ Aql	8.620	0.007	7.022	0.054	17.1399	0.0040	0.485	0.031	0.043	21	21
TT Aql	7.119	0.009	5.606	0.035	13.7539	0.0022	0.449	0.062	0.067	36	36
η Aql	3.894	0.006	2.999	0.025	7.1767	0.0004	0.332	0.042	0.022	46	46
SY Aur	9.073	0.008	7.806	0.028	10.1445	0.0017	0.293	0.036	0.026	22	22
RX Aur	7.677	0.006	6.591	0.021	11.6250	0.0016	0.278	0.029	0.031	27	27
RW Cam	8.672	0.008	7.027	0.024	16.4158	0.0019	0.341	0.048	0.056	49	49
RW Cas	9.243	0.013	7.709	0.028	14.7874	0.0022	0.481	0.075	0.139	50	50
SZ Cas	9.854	0.005	8.063	0.017	13.6396	0.0019	0.202	0.022	0.047	28	28
X Cyg	6.383	0.005	5.257	0.018	16.3860	0.0017	0.422	0.054	0.067	130	130
TX Cyg	9.513	0.009	7.135	0.033	14.7153	0.0018	0.450	0.066	0.060	45	45
W Gem	6.952	0.006	5.905	0.029	7.9133	0.0008	0.348	0.032	0.036	30	30
AA Gem	9.723	0.008	8.542	0.020	11.2935	0.0022	0.286	0.049	0.041	38	38
ζ Gem	3.919	0.005	3.049	0.020	10.1543	0.0019	0.220	0.029	0.026	34	34
Z Lac	8.412	0.008	7.125	0.029	10.8847	0.0016	0.386	0.060	0.030	38	38
T Mon	6.133	0.008	4.954	0.054	27.0055	0.0077	0.425	0.038	0.061	23	23
SV Mon	8.254	0.016	7.085	0.038	15.2476	0.0068	0.472	0.086	0.073	25	25
YO ph	6.169	0.004	4.483	0.017	17.1341	0.0029	0.216	0.023	0.025	39	39
SV Per	8.981	0.011	7.625	0.029	11.1274	0.0030	0.365	0.059	0.088	32	32
VX Per	9.306	0.010	7.836	0.025	10.8829	0.0030	0.296	0.057	0.069	31	31
RS Pup	6.993	0.011	5.419	0.047	41.5256	0.0224	0.435	0.051	0.053	21	21
AQ Pup	8.758	0.008	7.085	0.051	30.0124	0.0214	0.453	0.037	0.066	30	30
S Sge	5.619	0.005	4.743	0.021	8.3822	0.0008	0.322	0.034	0.032	43	43
W Sgr	4.668	0.006	3.849	0.025	7.5946	0.0004	0.337	0.037	0.027	39	39
X Sgr	4.564	0.005	3.670	0.020	7.0125	0.0004	0.276	0.030	0.042	44	44
WZ Sgr	8.011	0.009	6.536	0.037	21.8451	0.0057	0.476	0.059	0.078	37	37
RY Sco	8.015	0.007	6.225	0.035	20.3142	0.0055	0.375	0.039	0.044	29	29
Y Sct	9.621	0.008	7.697	0.022	10.3397	0.0017	0.315	0.058	0.045	46	46
Z Sct	9.581	0.006	8.069	0.027	12.9039	0.0014	0.403	0.040	0.057	50	50
RU Sct	9.475	0.012	7.297	0.039	19.6985	0.0035	0.476	0.085	0.089	31	31
SV Vul	7.206	0.008	5.622	0.029	44.9629	0.0203	0.412	0.046	0.049	34	34
LMC											
HV 874	14.454	0.026	12.6820	0.0000	0.467	0.110	0.000	10	0
HV 877	13.370	0.013	12.150	0.039	45.1135	0.0441	0.256	0.112	0.181	22	22
HV 879	13.334	0.016	12.309	0.058	36.8521	0.0195	0.457	0.094	0.101	14	14
HV 899	13.452	0.034	12.542	0.084	31.0559	0.0345	0.504	0.155	0.135	9	9
HV 900	12.876	0.037	11.826	0.098	47.2786	0.4734	0.452	0.120	0.102	7	7
HV 909	12.799	0.039	12.146	0.080	37.1650	0.7927	0.375	0.039	0.080	6	6
HV 953	12.312	0.018	11.386	0.034	47.9944	0.0252	0.393	0.118	0.100	24	24
HV 955	14.058	0.021	13.7320	0.0000	0.388	0.079	0.000	9	0
HV 971	14.497	0.029	9.2970	0.0000	0.244	0.162	0.000	9	0
HV 997	14.537	0.019	13.615	0.063	13.1465	0.0128	0.412	0.050	0.069	9	9
HV 1002	12.757	0.033	11.997	0.040	30.4797	0.0106	0.496	0.054	0.116	13	13
HV 1005	14.027	0.025	18.7100	0.0000	0.448	0.087	0.000	9	0
HV 1013	13.831	0.017	12.847	0.039	24.1730	0.0168	0.350	0.054	0.106	15	15
HV 1023	13.465	0.056	12.694	0.047	26.6587	0.0313	0.587	0.034	0.223	8	8
HV 2257	13.048	0.024	11.930	0.073	38.7990	0.1239	0.485	0.065	0.138	9	9
HV 2260	14.921	0.026	14.072	0.059	12.9882	0.0138	0.335	0.070	0.102	8	8
HV 2294	12.668	0.013	11.820	0.038	36.5471	0.0067	0.466	0.084	0.089	31	31
HV 2301	13.947	0.013	9.4990	0.0000	0.214	0.044	0.000	10	0
HV 2324	14.331	0.025	13.401	0.047	14.3916	0.0164	0.391	0.082	0.062	9	9
HV 2338	12.757	0.018	11.781	0.054	42.1609	0.0247	0.459	0.159	0.136	20	20
HV 2352	14.172	0.020	13.410	0.039	13.6201	0.0128	0.291	0.063	0.057	10	10
HV 2369	12.617	0.014	11.601	0.038	48.3646	0.0195	0.456	0.087	0.089	29	29
HV 2527	14.623	0.032	12.9480	0.0000	0.467	0.111	0.000	8	0
HV 2549	13.702	0.012	16.1970	0.0000	0.455	0.032	0.000	8	0
HV 2579	13.963	0.019	13.4310	0.0000	0.418	0.072	0.000	9	0
HV 2580	13.936	0.014	13.166	0.067	16.9511	0.0122	0.374	0.044	0.055	9	9
HV 2733	14.685	0.012	14.014	0.040	8.7269	0.0056	0.194	0.060	0.075	20	9
HV 2793	14.101	0.063	13.164	0.055	19.1840	0.0000	0.409	9.999	0.170	4	4
HV 2827	12.306	0.014	11.088	0.032	78.5562	0.2910	0.222	0.050	0.079	13	13
HV 2836	14.608	0.022	13.575	0.078	17.4706	0.0178	0.399	0.054	0.077	7	7

TABLE 4
(Continued)

Name	$\langle V \rangle$	σ	$\langle I \rangle$	σ	P	σ	A	σ_V	σ_I	n_V	n_I
HV 2854	14.633	0.010	13.928	0.063	8.6374	0.0036	0.242	0.042	0.055	17	6
HV 2864	14.628	0.020	13.820	0.064	10.9856	0.0052	0.385	0.058	0.069	10	10
HV 5497	11.916	0.007	10.753	0.016	99.7873	0.0932	0.226	0.067	0.059	62	30
HV 6105	14.917	0.026	10.4400	0.0000	0.299	0.086	0.000	8	0
HV 12700	14.849	0.008	14.037	0.032	8.1536	0.0028	0.231	0.035	0.053	17	8
HV 12815	13.451	0.018	12.484	0.050	26.1040	0.0095	0.448	0.118	0.092	14	14
HV 12816	14.516	0.012	13.873	0.040	9.1038	0.0041	0.305	0.064	0.051	22	11
HV 12823	14.569	0.020	8.3020	0.0000	0.371	0.070	0.000	9	0
SMC											
HV 817	13.864	0.007	13.062	0.031	18.8984	0.0044	0.365	0.098	0.077	52	34
HV 823	13.755	0.012	12.748	0.040	31.8877	0.0219	0.379	0.155	0.132	34	22
HV 824	12.370	0.006	11.395	0.027	65.8572	0.0128	0.365	0.061	0.085	62	34
HV 829	11.921	0.006	11.053	0.043	85.8151	0.0843	0.310	0.027	0.045	19	16
HV 834	12.180	0.006	11.303	0.019	73.3750	0.0243	0.251	0.053	0.077	62	31
HV 836	14.785	0.017	9.4034	0.0000	0.392	0.061	0.000	11	0
HV 837	13.241	0.010	12.327	0.044	42.6483	0.0133	0.391	0.054	0.052	23	20
HV 840	13.586	0.013	12.724	0.053	33.1165	0.0065	0.464	0.074	0.101	31	21
HV 843	15.037	0.020	13.871	0.052	14.7348	0.0344	0.330	0.059	0.262	11	3
HV 847	13.906	0.014	12.923	0.052	27.0935	0.0096	0.373	0.141	0.106	24	13
HV 854	14.245	0.030	15.9530	0.0000	0.396	0.144	0.000	11	0
HV 856	14.886	0.020	12.1553	0.0000	0.315	0.067	0.000	9	0
HV 857	14.429	0.015	11.9829	0.0000	0.388	0.044	0.000	8	0
HV 863	13.329	0.018	12.441	0.078	28.9201	0.0203	0.470	0.127	0.100	16	10
HV 865	13.112	0.008	12.367	0.072	33.3075	0.0119	0.450	0.025	0.153	15	10
HV 1326	14.838	0.023	13.7274	0.0000	0.382	0.107	0.000	12	0
HV 1334	14.874	0.027	9.4514	0.0000	0.406	0.126	0.000	12	0
HV 1338	15.103	0.019	8.4934	0.0000	0.387	0.061	0.000	11	0
HV 1342	14.207	0.008	13.493	0.022	17.9507	0.0076	0.210	0.051	0.061	37	25
HV 1365	15.015	0.007	14.260	0.057	12.4115	0.0016	0.304	0.081	0.139	48	16
HV 1695	14.723	0.018	13.803	0.035	14.5800	0.0091	0.297	0.083	0.065	18	18
HV 1744	14.559	0.025	13.774	0.056	12.6201	0.0078	0.331	0.114	0.072	16	8
HV 1787	14.278	0.012	13.507	0.108	16.1804	0.0105	0.350	0.040	0.077	13	5
HV 1873	14.847	0.023	14.039	0.173	12.9248	0.0063	0.402	0.061	0.108	10	2
HV 1877	13.146	0.012	12.177	0.060	49.5681	0.1012	0.226	0.032	0.092	12	12
HV 1884	14.435	0.019	13.521	0.056	18.0905	0.0165	0.260	0.190	0.120	21	8
HV 1954	13.864	0.018	13.083	0.049	16.6993	0.0127	0.310	0.099	0.079	16	9
HV 2017	14.699	0.017	13.938	0.071	11.4118	0.0016	0.334	0.141	0.066	31	7
HV 2052	14.241	0.021	12.5750	0.0000	0.361	0.073	0.000	9	0
HV 2063	14.749	0.014	11.6620	0.0000	0.313	0.046	0.000	9	0
HV 2064	13.722	0.014	12.741	0.059	33.6694	0.0121	0.416	0.069	0.070	22	12
HV 2087	15.199	0.024	9.1592	0.0000	0.296	0.079	0.000	9	0
HV 2088	14.663	0.010	13.700	0.057	14.5778	0.0013	0.374	0.058	0.089	24	9
HV 2103	15.132	0.019	8.9841	0.0000	0.303	0.065	0.000	9	0
HV 2189	14.487	0.011	13.4591	0.0000	0.344	0.033	0.000	9	0
HV 2195	13.009	0.012	12.127	0.038	41.9100	0.0325	0.359	0.066	0.094	22	14
HV 2202	14.408	0.014	13.643	0.058	13.1951	0.0022	0.311	0.063	0.055	21	6
HV 2205	13.999	0.024	13.196	0.099	25.5197	0.0163	0.459	0.139	0.143	17	9
HV 2209	13.553	0.006	12.871	0.028	22.6470	0.0045	0.293	0.031	0.064	23	15
HV 2225	14.781	0.014	13.954	0.032	13.1450	0.0053	0.279	0.042	0.107	15	8
HV 2227	14.801	0.008	13.873	0.030	12.4679	0.0018	0.310	0.038	0.081	23	11
HV 2230	14.658	0.020	13.797	0.049	12.5178	0.0069	0.279	0.049	0.030	7	4
HV 2231	13.508	0.020	12.591	0.084	36.6686	0.0244	0.385	0.150	0.084	16	10
HV 2233	13.874	0.010	13.180	0.066	15.1647	0.0023	0.443	0.025	0.058	8	5
HV 6320	14.832	0.013	14.133	0.041	10.1006	0.0026	0.417	0.097	0.075	35	24

in terms of Fourier expansions (e.g., Simon and Lee 1981; Moffett and Barnes 1985; Antonello and Poretti 1986; Peterson 1986). It is found empirically that the numerical values of the Fourier amplitudes and phases vary in a well-defined way with the period of the variable (the "Hertzsprung sequence," e.g., Ledoux and Walraven 1958). Therefore I have adopted Fourier decomposition with amplitudes and phases

that vary continuously with period as the means of encoding our template light curves:

$$V(t) = V_0 + \sum_{j=1}^n A_j \cos[j\omega(t) + \gamma_j],$$

$$I(t) = I_0 + \sum_{j=1}^n B_j \cos[j\omega(t) + \delta_j],$$

where $\omega(t) \equiv 2\pi\phi(t) = 2\pi[(t-t_0)/P]$.

I selected a sample of 114 Milky Way and Magellanic Cloud Cepheids having published photometry in the Johnson *V* and Kron-Cousins or Johnson *I* bandpasses. Only Cepheids with at least eight observations and lying in the period range $7 \leq P \leq 100$ days were considered, because shortward of 7 days, overtone and multiple-mode Cepheids complicate the picture, while longward of 100 days the period-luminosity relation is not well known, so these stars are of little interest for distance-scale studies. The actual stars employed in this analysis are listed in Table 3.

In the case of the Milky Way Cepheids studied by Moffett and Barnes (1980), it was necessary to convert the *I*-band magnitudes from the Johnson system to that of Kron and Cousins; this was done employing a relation based on results reported by Taylor (1986; see his Table 4):

$$(R-I)_{KC} = 0.873[(R-I)_{MB} - 0.008] + 0.049,$$

$$(V-R)_{KC} = 0.714(V-R)_{MB} - 0.029.$$

Neglect of Taylor's right-ascension-dependent corrections will introduce positive and negative systematic errors of less than 0.01 mag in absolute value; furthermore, since these will affect solely the average magnitudes of the stars, not the amplitude or shape of the light curves, they are irrelevant for the present application.

I adopted a five-term expansion for the *V* light curves, and four terms for *I*; higher-order models did not materially improve the fit. Specifically, for the *i*th star:

$$V_i(t) = V_{0,i} + A_i [\cos \omega + \alpha_2 \cos(2\omega + \gamma_2) + \alpha_3 \cos(3\omega + \gamma_3) + \alpha_4 \cos(4\omega + \gamma_4) + \alpha_5 \cos(5\omega + \gamma_5)],$$

$$I_i(t) = I_{0,i} + A_i [\beta_1 \cos(\omega + \delta_1) + \beta_2 \cos(2\omega + \delta_2) + \beta_3 \cos(3\omega + \delta_3) + \beta_4 \cos(4\omega + \delta_4)],$$

where $\omega = \omega(t) = 2\pi\phi_i(t) = 2\pi[(t-t_{0,i})/P_i]$. Simultaneous least-squares solutions of these equations determined the coefficients of α , β , γ , and δ that define the template light curves, as well as the five fundamental quantities for each Cepheid: V_0 , I_0 , A , P , and t_0 . (The published period for each variable star was adopted as a starting guess; the least-squares solution was allowed to try to improve upon it.) This approach differs slightly from that taken by most previous studies, where more typically a low-order Fourier expansion was fitted to individual light curves, and then the behavior of the various components as a function of period was investigated (e.g., Simon and Lee 1981). Here, building on the work of these previous investigators, I have begun with the assumption that all amplitudes and phases are continuous functions of period, with one exception: it has been known for some time that the phases of the second and third Fourier components go through an abrupt change at a period of around 10 days. My own preliminary analysis, which was based solely on Cepheids with periods greater than 10 days, showed that the amplitudes of both the $\cos(2\omega)$ and $\cos(3\omega)$ terms appeared to extrapolate to zero at almost exactly the same period just short of 10 days. Of course, if the amplitude of a Fourier component is zero, its phase is undefined, while for Cepheids with periods near the crossover

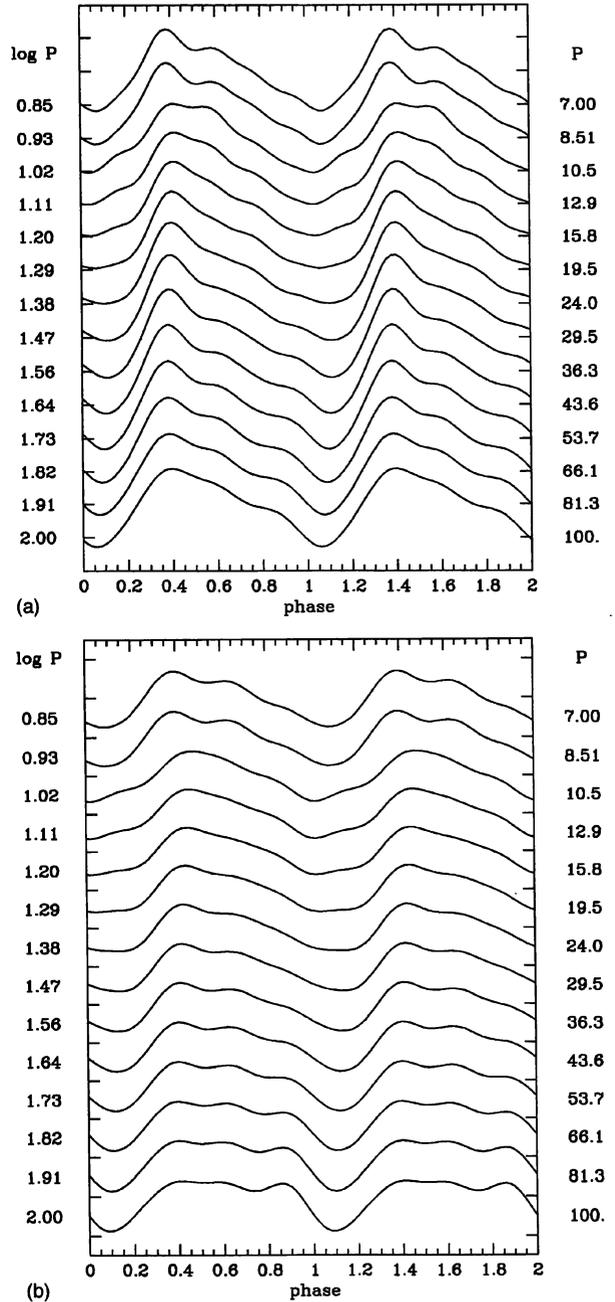


FIG. 14—The actual template light curves obtained by the present Fourier decomposition of data for Milky Way and Magellanic Cloud Cepheids. The model light curves are shown for periods in the range 7–100 days, with an increment $\Delta \log(P) = 0.0888$ between models. (a) The *V*-band templates; (b) the *I*-band templates.

period, the phase of each component is extremely poorly constrained. When the amplitudes of the second and third Fourier components become significant again below the crossover period, it is found that a large, abrupt change in their phase has occurred (see, e.g., Simon and Lee's Figs. 1 and 4). Therefore my model light curves incorporate a discontinuity in the phase of the second and third Fourier components at $\log(P) = 0.94, P = 8.71$ days. By performing simul-

taneous Fourier fits to all stars at once, rather than individually, I am able to force continuity in the *amplitudes* of the second and third Fourier components as they pass through zero, while the phases of these components for periods where their amplitudes are approaching zero are also forced to be continuous and well defined. The amplitudes of the fourth and fifth Fourier components do not approach zero anywhere in the range $7 < P < 100$ days, ($[0.84 < \log(P) < 2.0]$), so a single continuously varying relation was imposed for both the amplitude and phase of these components, as well as for the first Fourier component of the *I*-band relation.

With $X \equiv \log(P) - 1$, the least-squares solution of these equations gives the light-curve parameters listed in Table 3 and the stellar parameters listed in Table 4. The root-mean-square deviation of 2751 individual *V*-band observations from 114 template fits was 0.045 mag; the rms deviation of 2002 individual *I*-band observations from 92 template fits was 0.054 mag. Figure 14(a) illustrates the resulting template light curves for the period range $7 < P < 100$ days; the vertical scale is arbitrary (i.e., these templates are taken to apply to Cepheids of all amplitudes), and successive curves are separated by $\Delta \log(P) = 0.0888$. Figure 14(b) shows the same for the *I*-band light curves.

Figures 15–18 show the residuals of individual observations from the best-fitting template for each star as a function of phase for various subsamples. In Fig. 15 the *V*-band residuals have been plotted for the present sample of Cepheids broken down into four period ranges, while Fig. 16 shows the same for the *I*-band residuals. While some systematic fillips appear at some phases for some periods—most notably near just before maximum light for Cepheids with $10 < P < 20$ days—these are comparatively minor and of the same order of magnitude as the intrinsic dispersion. Quite a

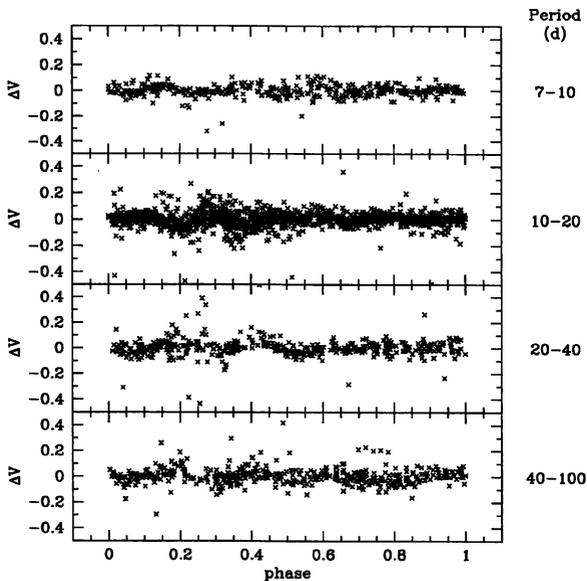


FIG. 15—Individual *V*-band fitting residuals vs. phase for all Galactic and Magellanic Cloud Cepheids used in the decomposition, divided into four intervals of period.

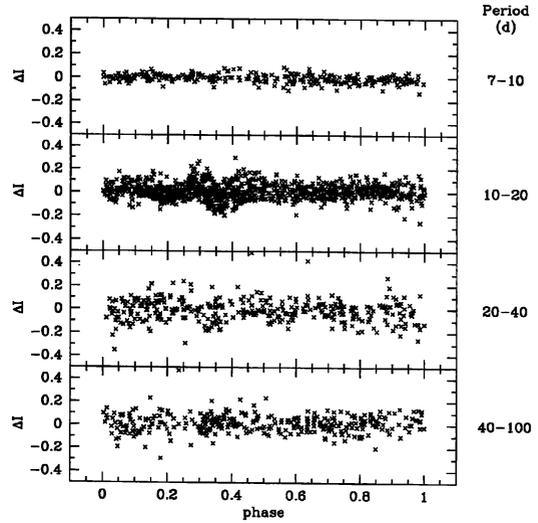


FIG. 16—Individual *I*-band fitting residuals vs. phase for all Galactic and Magellanic Cloud Cepheids used in the decomposition, divided into four intervals of period.

large number of additional Fourier components would have to be added to the model to flatten the residual plots completely. Such wiggles as do occur will average out over the full cycle, as is required by the least-squares formulation, and should have minimal deleterious effect on the estimation of periods and mean magnitudes. Together Figs. 15 and 16 suggest that the fitting residuals owe as much or more to random observational errors in the published data, or random differences in the light curves of different Cepheids with the same period, than they do to systematic inadequacy of the templates.

Figure 17 shows the *V*-band residuals broken down by home galaxy; it seems that the same templates apply equally well to the Milky Way and the two Magellanic Clouds. The slight bobble in the Milky Way residuals near phase 0.27 may be due as much to the slightly different distribution of

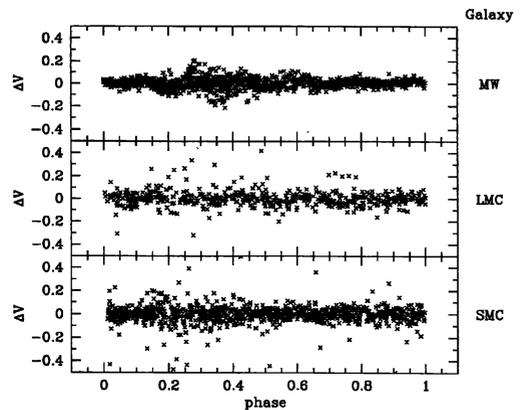


FIG. 17—Individual *V*-band fitting residuals vs. phase for all Galactic and Magellanic Cloud Cepheids used in the decomposition, divided according to parent galaxy.

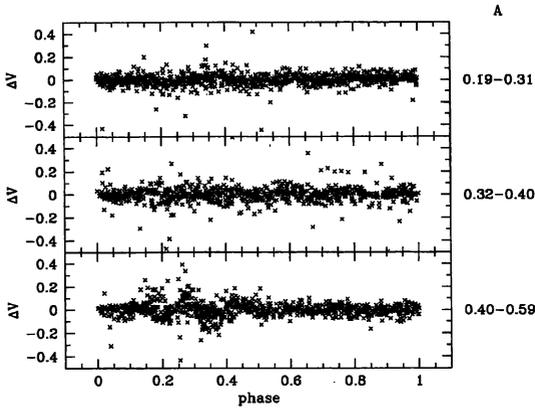


FIG. 18—Individual V-band fitting residuals vs. phase for all Galactic and Magellanic Cloud Cepheids used in the decomposition, divided into three intervals of amplitude.

Cepheids over period in the different samples as to any intrinsic difference at fixed period. Figure 18 shows the same V-band residuals divided according to the first-harmonic visual semi-amplitude, A ; dividing points were chosen to place 36 variables in each subsample. Again, the structure seen in the diagram for the variables with the largest amplitude may be due to nonuniform filling of the period–amplitude plane (see below), rather than to any true amplitude dependence of the pulsation cycle. Even if there is a real amplitude effect, it is small enough that it should not jeopardize a useful employment of these templates.

Figures 19–21 show the raw data and the fitted template light curves for three stars showing, respectively, typical,

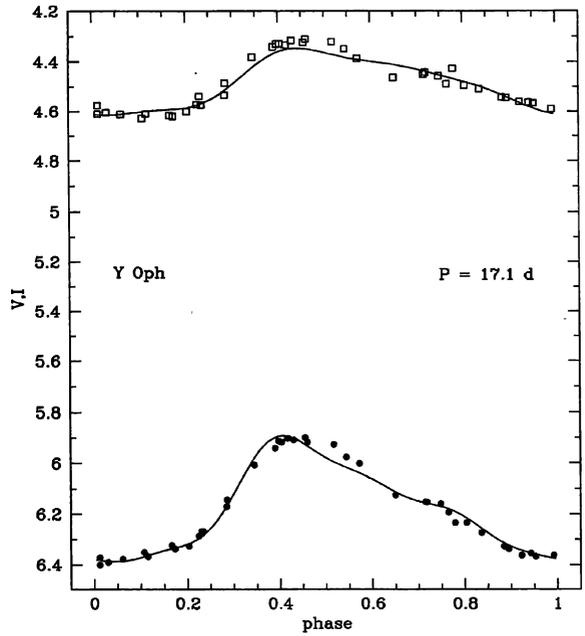


FIG. 20—Actual data and fitted template light curves for the Galactic variable Y Oph, which illustrates fitting errors for my best case. Filled circles are V-band data and empty squares are I-band data.

minimum, and maximum root-mean-square residuals. RW Cam had normalized rms residuals of 0.048 and 0.056 mag in V and I, respectively—as close to the overall sample mean as any star with a reasonable number of observations (49 in each bandpass). Y Oph had rms residuals of 0.023 and

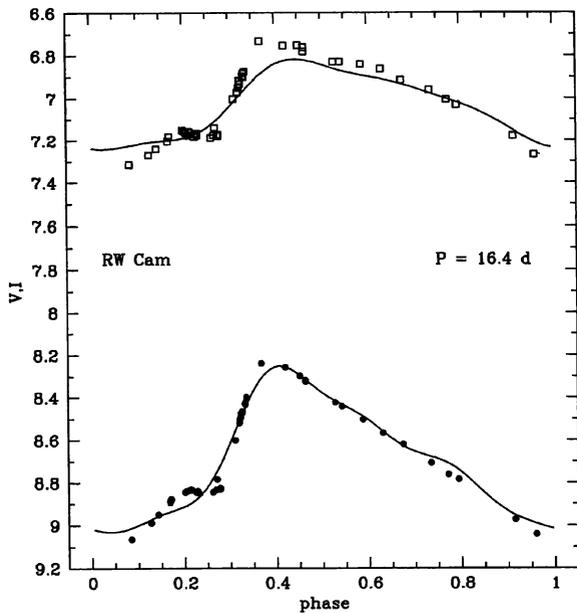


FIG. 19—Actual data and fitted template light curves for the Galactic variable RW Cam, which illustrates fitting errors for a typical case. Filled circles are V-band data and empty squares are I-band data.

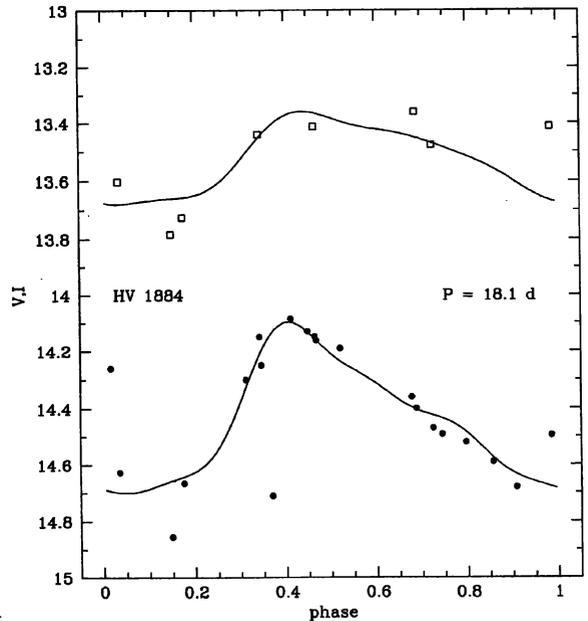


FIG. 21—Actual data and fitted template light curves for the SMC variable HV 1884, which illustrates fitting errors for my worst case. Filled circles are V-band data and empty squares are I-band data.

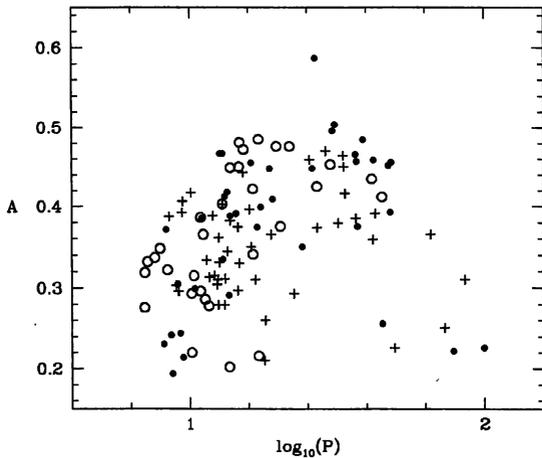


FIG. 22—The period–amplitude relation for all Galactic and Magellanic Cloud Cepheids employed in the present template determination. Large empty circles represent Cepheids in the Galaxy, smaller filled circles represent those in the LMC, while plusses represent those in the SMC.

0.025 mag in the two bandpasses and was the best-fit star, while the worst-fit star, the SMC variable HV 1884, had rms residuals of 0.19 and 0.12 mag; Fig. 21 suggests that this is due mostly to large observational errors in a few data points.

Finally, Fig. 22 shows the period–amplitude plane for the Milky Way and Magellanic Cloud Cepheids used to define the template relations. Here empty circles represent variables in the Galaxy, filled circles are for the LMC, and “+” signs are for the SMC. Besides serving as a basis of reference for the distribution of periods and amplitudes found in other galaxies, this diagram has the noteworthy feature that there appears to be a deficiency of long-period, low-amplitude Cepheids in the Galactic sample. Quite possibly this is a

selection effect: it could be that the Magellanic Clouds, with their comparatively small angular extent, have been more effectively searched for long-period, low-amplitude variables than the plane of the Galaxy, which covers almost a third of the sky and is beset with extinction problems. The preponderance of Galactic Cepheids with largish amplitudes and periods in the range $10 < P < 25$ days may account in part or in full for the apparent systematic departures seen in the top panel of Fig. 17 and the bottom panel of Fig. 18.

REFERENCES

- Antonello, E., and Poretti, E. 1986, *A&A*, 169, 149
 Freedman, W. L., et al. 1994, *ApJ*, 427, 628
 Hughes S. M. G., et al. 1994, *ApJ*, 428, 143
 Lafler, J., and Kinman, T. D. 1965, *ApJS*, 11, 216
 Ledoux, P., and Walraven, T. 1958, *Handbuch der Phys.*, 51, 353
 Moffett, T. M., and Barnes, T. G. III 1980, *ApJS*, 44, 427
 Moffett, T. M., and Barnes, T. G. III 1985, *ApJS*, 58, 843
 Peterson, J. O. 1986, *A&A*, 170, 59
 Pierce, M. J., Welch, D. L., McClure, R. D., van den Bergh, S., Racine, R., and Stetson, P. B. 1994, *Nature*, 371, 385
 Saha, A., Labhardt, L., Schwengeler, H., Macchetto, F. D., Panagia, N., Sandage, A., and Tammann, G. A. 1994, *ApJ*, 425, 14
 Sandage, A., Saha, A., Tammann, G. A., Panagia, N., and Macchetto, D. 1992, *ApJ*, 401, L7
 Simon, N. R., and Lee, A. S. 1981, *ApJ*, 248, 291
 Stellingwerf, R. F. 1978, *ApJ*, 224, 953
 Stetson, P. B. 1987, *PASP*, 99, 191
 Stetson, P. B. 1989, in *Image and Data Processing: Interstellar Dust*, V Escola Avançada de Astrofísica, ed. B. Barbuy, E. Janot-Pacheco, A. M. Magalhães, and S. M. Viegas, Universidade de São Paulo, p. 1
 Stetson, P. B. 1994, *PASP*, 106, 250
 Welch, D. L., and Stetson, P. B. 1993, *AJ*, 105, 1813
 Tanvir, N. R., Shanks, T., Ferguson, H. C., and Robinson, D. R. T. 1995, *Nature*, 377, 27
 Taylor, B. J. 1986, *ApJS*, 60, 577