The Combination of Spectroscopic and Astrometric Data in the Analysis of Orbital Motions in Binary Star Systems

CHRISTOPHER MORBEY[1]

Dominion Astrophysical Observatory, Herzberg Institute of Astrophysics, National Research Council of Canada, Victoria B.C., Canada

ABSTRACT: Even though various types of data for binary stars are being obtained in unprecedented quantities there is still a lack of completeness. Because the reduction of combinations of these data adds to the fundamental concepts in astronomy it is advantageous to have available a general self–consistent method whereby all the elements of the orbits of a binary or multiple system may be solved simultaneously. Such a method is reviewed in this paper and an emphasis is placed on the specification of meaningful errors attached to the orbital elements.

1. INTRODUCTION

Astronomy, I would say, is a well–balanced discipline. A reasonably happy mixture of theory, observation, speculation and experimentation by computer–modeling would seem to have kept the oldest science alive and well through good times and bad. Perhaps more fundamental for the sustaining of astronomy today is the drive of curiosity (public pressure, if you like) which nearly every human of young spirit possesses. Luckily for us, in this instance the media have understood the general keen interest in astronomy and they have tried to keep the population informed. It is incumbent on us, therefore, and in particular, I think, those who would study binary and multiple systems, to gather as much data as possible and to understand it as thoroughly as possible. Only recently, Batten & Fletcher (1989) show that the degree of completeness of our knowledge of spectroscopic binaries is seriously lacking. In spite of the new catalogues in recent years I would venture to guess that the same predicament exists for astrometric binaries and even more so for multiple systems. To be commended are those individuals who, year after year, publish observations and orbits in increasing numbers. But I believe we can't let up. In a nutshell, we've got to get the most out of our data since it is our work which provides a foundation for studies in star formation and evolution, the mass/luminosity relation, and galactic structure and dynamics. It is equally important not to derive more from the data than is warranted.

My assigned topic is the combination of different data types, especially radial velocities and astrometric data, in the analysis of orbital motions in binary star systems. One possible approach would be to enumerate all published methods and evaluate the relative advantages and disadvantages of each. Although that approach might be reasonably objective it would be rather involved and would require an amount of effort which would easily be a candidate for submission to the supplement of some Journal and involve more than one hundred

---

[1]This paper was read in Dr. Morbey's absence by Dr. Colin Scarfe

pages. Missed would be the many private methods which computers use but which have never been published. The approach I will take is a far cry easier since my ideas will be discussed. Of particular interest to me (as I've alluded to) is to get the very most possible out of the data but no more. Significance or confidence intervals, variances, constraints, weights, and systematic and internal errors are quantities which play a very important role in the kind of analyses I will summarize. In the first instance there is the general method which can be applied to any kind of binary system or even multiple system. Then there are the errors in the derived parameters and I shall concentrate on that aspect

2.   THE METHOD

Formalism describing a synthesis of the solutions of spectroscopic and visual binary orbits was described by Morbey (1975). Some years earlier Alan Batten and I had been working on the triple system ADS 11060 and we wondered whether or not the visual data and spectroscopic data could be combined in such a way as to predict more reliably the time of periastron passage of the visual double. Alan was also working on ADS 8189 at the time and was having some difficulty reconciling the visual and spectroscopic orbital solutions. Discussions involving these problems and some queries from the late W. S. Finsen led to the development of the method.

It is probably best at the outset to underline the fact that there is no real difficulty in combining different types of data with different units in a general least-squares problem of any sort. Once this fact is understood the rest should follow without undue difficulty. Another point is that nearly anybody's brand of least-squares is sufficient to solve the problem. For in depth studies of the least-squares reduction methods you could look up a large number of references. In particular, and rather pertinent to the problem at hand is the reference by Eichhorn & Xu (1990) and earlier works by Jeffreys (1980, 1981). There are numerous texts which go into great detail for many aspects of the problem. Notably here is the book *Numerical Recipes: The Art of Scientific Computing* by Press *et al.* (1986) or *Data Reduction and Error Analysis for the Physical Sciences* by Bevington (1969).

So, what DO you do to combine data of different types in a least-squares solution? Some critics would suggest that you can't combine apples and oranges. My trite reply might be that my wife and I do it all the time and when put through the juicer the mixture is really quite tasty. In the case of least-squares the result is nearly as delightful; you only have to normalize the squared differences, that is, the $(O - C)$'s, by the variances and dimensionless quantities result; the sum of all these is the so-called "chi-square". So whether you pick radial velocity, radial velocity difference, ρ, θ, or intensity, it matters not... just as long as you divide by their variances. The problem then reduces to a determination of what the variances are and how the various quantities are weighted with respect to one another. Problems of adding or not adding 180 degrees to ω, Ω, or the position angles are not difficult... they're just a bit confusing.

In mathematical notation the combination of the different groups of data into a simultaneous solution is easily accomplished by formulating the objective function, χ^2, in the following way.

$$\chi^2 = \sum_{j=1}^{M} \sum_{i=1}^{N} \left[\frac{O_{ij} - C_{ij}}{\sigma_{ij}} \right]^2 \tag{1}$$

The O_{ij} are the observables, each designated i (total N) in a group j (total M). C_{ij} refers to the corresponding calculated values which, in the case here, are found from the usual equations used to determine the elements of visual and spectroscopic binary orbits. The σ_{ij} are the variances. In order to obtain the best or most realistic values for the C_{ij} the χ^2 function must be minimized since this makes the average deviation of the observed quantities from the calculated quantities as small as possible. And remember, the residuals have been reduced to dimensionless quantities with the variance scale factors so there is no problem with the combination of the data groups having different dimensions. Both spectroscopic and visual orbital elements are contained within the equations which are used to compute the C_{ij}. For this reason the solution can be called simultaneous. We also call the solution self-consistent because the different χ^2 sums can be relatively weighted. Changes to the elements can be described by

$$e = (\alpha^T W \alpha)^{-1} \alpha^T W O \quad, \tag{2}$$

where W is the matrix of relative weights and α is the design matrix made up of the appropriate derivatives. And because we are dealing with dimensionless quantities the matrix elements referring to the different data types can be summed.

This is the simple minded approach, and if you are so inclined there is little complexity in working out a fortran program to do the work. It really isn't necessary to do that of course because you'll find all sorts of listings for least-squares programs in the literature. One very useful reference and probably the easiest to implement is the book by Bevington (1969). His routine called CURFIT works with spectroscopic and/or visual binary problems very well. It uses the method of damped least–squares originally devised by Levenberg (1941) and popularized by Marquardt (1963). In that technique, the best features of the gradient search method and the method of linearizing the fitting function are combined; the diagonal terms of the design matrix are augmented by some factor which controls an interpolation of the algorithm between the two method extremes. We note that the inverse of the design matrix is called the error matrix and contains estimates of both the variances and covariances of the orbital parameters. This is just basic to the usual method of least squares.

We have left out some embellishments here. If you look at the work by Morbey (1975) or by Eichhorn & Xu (1990) you will note that there is provision to introduce constraints. Those of you who are familiar with solving any kind of astronomical problem know that the introduction of constraints can be a real headache. Even the relative constraints we call weights can add a whole new dimension to the simple solution of some problem. However, we are determined to extract the most out of our data so we've got to give the solution mechanism every ounce of information we can muster. If we know a certain parameter must have a certain value we introduce an equality constraint. If we know the parameter must take on values only within a certain range then we have an inequality constraint. Because our problem is rather non-linear we are now

involved with non-linear programming. With this kind of expertise you might want to look at a career in economics if you should not be successful at binary star data. Once you embark in economic modeling you will never have to worry about being wrong again because you always will be. Seriously, there are many tricks of the trade here, and if you are just starting into the exciting realm of solving binary systems by combining data of different types I would suggest that you start with something simple like CURFIT and once that is mastered go on from there. There are actually quite a few simple additions you can make to CURFIT or similar routines which can assist in solving the kinds of problems we are dealing with. One particularly useful technique which can be used to satisfy the inequality constraint of some parameter is to fix the parameter at the boundary value if it is surpassed at an iteration and then reiterate, letting the parameter vary only after the first iteration. Note that the objective or merit function can be larger once the iterations start again. This is a way of allowing a new region of the hypersurface to be accessed. Recall that the sum of the squares of the residuals divided by the variances forms the so-called hypersurface which has dimensions equal to the number of parameters. Because of the non-linear nature of the problem and observational error the surface is anything but globally smooth. Usually the surface has many local minima, and the trick is to find the global minimum which satisfies all the constraints. Interdependencies among the parameters which are intrinsic or which are introduced because of the scatter in the various types of data only serve to cause difficulty for the mechanism used to locate the global minimum. They also confuse the interpretation of errors attached to the parameters.

In the last few years a very powerful technique for solving large non-linear problems has emerged. It is probably 'over–kill' for solving binary orbits because we are only interested in less than a dozen orbital parameters. But it would be rather interesting to see how successful it would be should some enterprising person care to do the analysis. The technique is called the 'simulated annealing method', and in problems which involve many parameters a global minimum can usually be assured. Essentially, an analogy to the physical process of crystallization is made. If some appropriate material at high temperature is slowly cooled, a low energy state can be achieved where the material crystallizes. Fast cooling or quenching does not result in a stable lowest energy state. A direct comparison can be made to the iterative improvement of complicated non-linear least-squares solutions, and the merit function or objective function takes on the role of energy. Small changes in parameters make small changes in energy or magnitude of the merit function. The real difference in this method over others is that while the algorithm implements changes in the parameters which yield a smaller merit function, it also accepts changes in the parameters which yield a larger merit function. This is done on a statistical basis. The probability that a configuration is accepted is the Boltzmann probability factor:

$$\exp(-E/kT) \ ,$$

where E refers to the energy or merit function, k is Boltzmann's constant and T is the temperature. If you are interested in the application of this method a useful reference is an article in Science by Kirkpatrick *et al.* (1983). Another is *The Art of Scientific Computing* (Press *et al.* 1986) mentioned earlier.

3. PRELIMINARY ESTIMATES OF THE ORBITAL ELEMENTS

Since we are dealing with a linearized non-linear least-squares problem, we are actually solving for the changes to the parameters and so we need some initial estimates. Before a solution is attempted or even before a data set is complete it is necessary to have a good idea of the period. This is usually the most difficult parameter to initialize. I won't go into any detail on period–finding methods since most everyone seems to be happy with the method they presently use. Let me just summarize the method that I use in case there are those who might like to try something different. Also in this kind of analysis the error is vitally important. We like to know what the chances are of extracting the same period from random data. More detail can be found in the paper by Morbey & Griffin (1987). Basically, there are three parts to the selection process:

1. the orbit curve is approximated by an empirical monotonic function;
2. a least-scatter selection criterion determines which periods yield arrangements of data which look 'best';
3. a bootstrap statistical comparison estimates the chance of finding a period of equal or better quality in random data.

Suppose we are looking for a period in a number of observed velocities. The variance of the velocities defines the standard of comparison with which trial orbits fit the velocities arranged at phases appropriate to a particular period. Then there is the quantity which describes the fit or gives the 'quality' of a trial period; it is the factor by which the variance from the empirical monotonic function is smaller than the standard variance. Lower qualities are attached to arrangements of velocity versus phase which do not follow the empirical function very well. What is this empirical function? If the average minimum velocity is constrained to occur at zero phase then it is defined by two best-fitting monotonic functions on either side of the phase of the average maximum velocity. The form of these functions is given by

$$x^n + y^n = 1 \ , \tag{3}$$

where the n is determined by the best fit. In practice, the fitting can be done before the trial periods are checked because the data are reduced into a 10×11 grid and there are only so many possibilities. To assess the absolute quality of a period we compute the chance that it could be detected in random data. By random data I mean data which is made up by randomly reassigning the observed velocities among the actual epochs of observation. In this way, period aliases in the original data are maintained in the random data.

Now that the period is isolated, what next? For the spectrographic elements we can simply choose as starting elements for the least–squares the average of the velocities for V_o, half the difference of the maximum and minimum velocities for K, 0.3 for e, 0 for ω, and the time of maximum velocity for T. These starting values are usually sufficient for most systems. For the initial visual elements, one can make some quick graphical plots or just make a few sensible guesses. Computer time is usually very cheap, and a dozen trials can be made even before you would have been able to find some graph paper.

4.   ERRORS IN THE ORBITAL ELEMENTS

Note that the elements in equation (2) are determined without knowledge, *a priori*, of the variances of the observables. As a summary, let's look at a few more equations which give us the errors in the parameters if the errors on the observables are gaussian. The covariance of e is written

$$covar(e) = \sigma^2(\alpha^T W \alpha)^{-1} \quad , \tag{4}$$

where σ^2 is the scale factor between the variances of the observables and the relative weights

$$var(O) = \sigma^2/W \quad . \tag{5}$$

The errors of the orbital elements are then

$$covar(e) = (\alpha^T V \alpha)^{-1} \quad , \tag{6}$$

where V is the matrix containing the inverses of the covariances. If the variances are not known they must be estimated before the errors of the elements can be computed. Supposing there are different kinds or groups of observables there are two ways to determine the variances:

1. Use internal errors.
2. Suppose weights for the observables such that the derived variance of the fit obtained from the sum of squared residuals is what might be expected.

The relative weights between the groups are not arbitrary but are selected iteratively such that the resultant variance of the fit of that group compares favourably with what is expected (χ^2 test). If the comparison is not favourable, it means that there are some systematic deviations not accounted for in the equations relating the orbital elements to the calculated values.

Normally, attention is only given to the diagonal terms (variances) of the covariance matrix, the square–roots designating the standard deviations of the orbital elements. However, the off-diagonal matrix elements are rarely zero and this means that some of the orbital elements are correlated with one another. The variances, then, are only part of the designation of errors. Although the covariances are seldom taken into account they really ought to be since a full knowledge of the deviations of a variable usually depends on the deviations of other variables. Eichhorn & Xu (1990) in their discussion of least–squares solutions to visual binaries point out the pitfalls of fixing particular orbital elements in order to obtain a better conditioned solution: "If the eliminated unknown was strongly correlated with some of the other, retained unknowns, the elimination of the correlation will have brought about smaller formal standard errors."

If we are to extract the most from our data, but no more, it is most important that the errors attached to the orbital elements are known precisely. Normally, as we have discussed, the errors are derived from the covariance matrix of the elements. In simple least–squares problems involving only one type of data, say radial velocities, we can assume sometimes that the observations

all have identical variance or equal weights, if you like. The elements can be determined without any knowledge of the variances but when it comes time to determine the errors of the elements we have to supply an estimate of the standard deviation of the residuals, and this is found directly from the observed values and the calculated best–fit values. The product of this standard deviation and the diagonal elements of the covariance matrix then yields the standard errors of the orbital parameters. In the case of solutions which combine various data types this really isn't possible because the standard deviations of the residuals have different units. To overcome this problem it is necessary to know, *a priori*, what the variances or standard deviations of the observed quantities are. They must be determined by obtaining the so–called internal errors of measurement. As we've mentioned, the only way around this procedure is to assume relative weights and iterate until the solved standard deviations of the fit turn out to be what would be expected. Of course, if this is done, you forego an independent evaluation of 'goodness of fit'. Visual observers with much experience can tell you quite accurately what they expect their errors of measurement to be, but let's be careful here. An objection to a general simultaneous solution of orbital elements by the method of least-squares is that you can't adjust the weights of the observations as you proceed. If you read older papers on the solutions of visual orbits you definitely get the impression that the solutions were 'moulded into shape' almost like a potter does with her clay. I would venture to guess that the great experience of these observers would be a satisfactory guide for the determination of definitive solutions, but in this day and age we are all too impatient for that kind of experience. And we've been taught over and over again that solutions should be as objective as possible. With the speed and capability of computers increasing daily there is no excuse to weight data according to how they fit a particular solution.

Once the least–squares procedure settles on some preliminary elements it is necessary to revisit the question of errors in the observations and how they propagate to errors in the elements. As I mentioned earlier it is necessary to make accurate assessments of the standard deviations of the observed quantities. This really is a separate subject, and we won't go into it here. We are nearly finished when the standard deviations which are determined from the least–squares fit match the original assessments. There will be times when such a match does not occur, and the question of reconciliation becomes paramount. Alan Batten at IAU Colloquium 62 (1981) had something to say about that. Obviously, a thorough knowledge of the internal errors of observation is most important.

It is of interest to look briefly at two data sets containing both spectroscopic and visual observations. We will, in particular, be looking out for correlations in the orbital elements rather than the actual values themselves. The data for Burnham 1163 were originally used to develop the the combined solution method whereas those for Struve 2173 were used just recently to determine its orbital elements completely.

First of all, we examine the orbital parameters of β 1163 as they have been determined with the very early visual data and the spectroscopic data described by Fletcher (1972). Normally distributed observation errors were superimposed on these data, and the solution was computed 100 times. Figure 1 shows the results of these calculations; each parameter plotted as a function of the other.
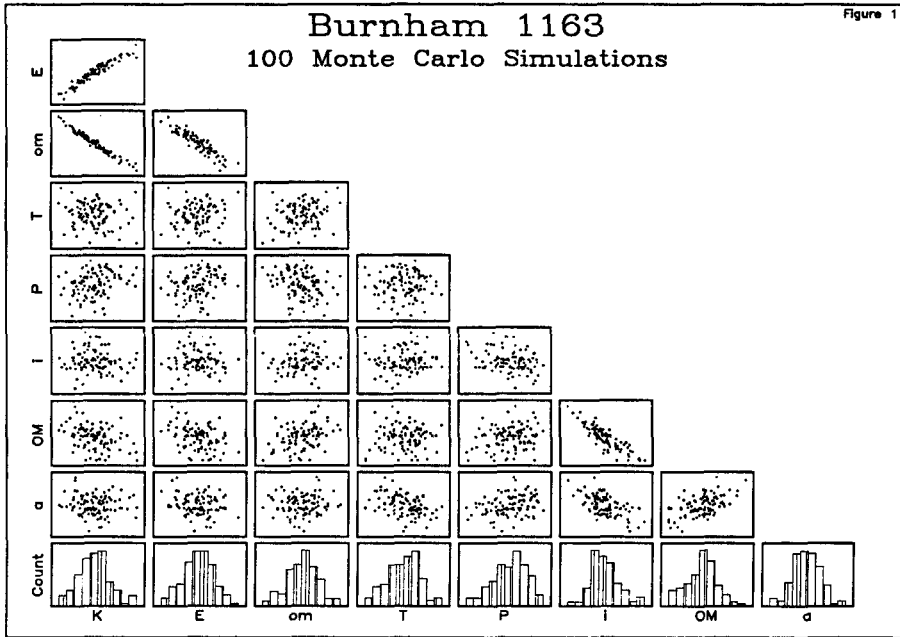
FIGURE 1.   Monte Carlo simulations of ADS 1123 = β 1163.

What is immediately obvious are the correlations among K, ω, and e. Next are correlations among i, Ω, and a. If any of these correlating elements are held constant the resulting errors of the elements they were correlated with are substantially reduced. This is just an artifact of doing something quite illegal in least–squares solutions, i.e. specifying a theoretical error in one parameter when another with which it is correlated is held constant. The data for β 1163 are given by Fletcher (1973) and Morbey (1975). Note that the visual data here are plotted on the 16–year period but they actually span a number of decades. As you can see, the radial velocities cover just one side of periastron; the velocities are relative and there is no minimum velocity difference defined. It is clear that random data which point to solutions which tend to have a reduced value of K have smaller eccentricities and larger oms since the curves are flatter. If you think about the correlations of i, Ω, and a for a while it becomes clear that they too result from freedom in the orbit shape–defining visual parameters. The point is that because the coverage of the orbit is not complete for all data types correlations develop. In these cases it is particularly important to analyse the errors properly.

Before leaving the discussion of these data let me also point out another pitfall which is a little more subtle. The random data sets produce standard deviations of the parameters which are nearly identical to the formal standard deviations derived from the covariance matrix. However, if instead of normally distributed errors we subject the data to uniformly distributed errors, the standard deviations become quite different and are usually substantially smaller. In
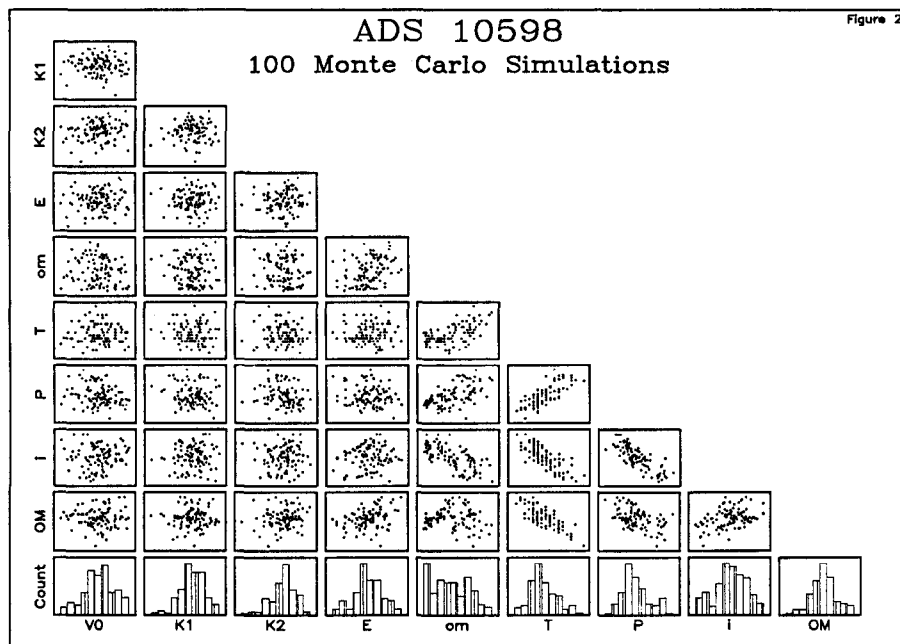
FIGURE 2. Monte Carlo simulations of ADS 10598 = Σ 2173.

sparse data sets the distributions tend to have outliers and although a uniform distribution may be a bit of an exaggeration it clearly shows what happens. A better determination of the form of distributions within the data types and use of these in the Monte Carlo simulations would then lead to more realistic error estimates.

Figure 2 shows the results for similar Monte Carlo estimates for Σ 2173. Coverage of the visual and spectroscopic data types in this case is much more complete but there are still correlations mainly involving T. If you are familiar with the data a little reflection shows why this is the case. Since the eccentricity is only 0.17, the radial velocity curves are reasonably sinusoidal. Moreover, the individual velocities of each component can only be determined when the lines have doubled, that is, near maximum and minimum velocity. For this reason T and ω are less precisely known and the resulting freedom in the Monte Carlo simulations causes correlations, mainly with P, i, and Ω. Again, the rule is: if there are correlations, don't fix one parameter and expect the errors of the others to be unaffected. If you do, in the case of this system, the errors can drop to only 10% of their real values.

5. CONCLUSION

I have summarized a self-consistent general method of solving the orbital elements from different types of data for binary star systems. Only ordinary χ^2 minimization techniques have been used but the various matrix elements sum

over different data types. Although I didn't refer to multiple systems, the same method can be used. The application of the least–squares method to the combined solution of binary orbits is really not all that complicated, and the elements can be obtained quickly if the data are sufficient. What takes more effort is a good understanding of the errors in the elements. Since most of us now have a substantial amount of computing power available, even Monte Carlo statistics should pose no real problem. The whole process is really not a 'sausage machine' method since most systems will have their own peculiar distribution of errors. I would advocate, therefore, that those who have found great satisfaction in computing the orbits of binary and multiple star system increase their satisfaction by including a full discussion of the errors of observation and the effects of their propagation through the calculations to the elements themselves. The tools are available, and I think we owe it to posterity.

6.   ACKNOWLEDGEMENTS

I would be remiss if I didn't bring attention to the fact that I have benefitted substantially from discussions with Alan Batten, Murray Fletcher, and Colin Scarfe over the years. Colin has graciously offered to present this paper, but I can't hold him responsible to answer all questions as I might. To save him from that you may reach me with email at morbey@dao.nrc.ca.

7.   REFERENCES

Batten A.H. 1981, IAU Colloquium 62, 271
Batten A.H. & Fletcher, J.M. 1989, *The Observatory*, **109**, 186
Batten A.H., Fletcher, J.M., Hill, G., Lu Wenxian, & Morbey, C.L. 1991, *PASP*, **103**, 294
Bevington, P.R. 1969, *Data Reduction and Error Analysis for the Physical Sciences*, (New York, McGraw-Hill)
Donaldson J.R. & Schnabel, R.B. 1987, *Technometrics*, **29**, 67
Eichhorn, H.K. & Xu, Yu-lin, 1990, *ApJ*, **358**, 575
Fletcher, J.M. 1973, *JRASC*, **67**, 255
Jeffreys, W.H. 1980, *AJ*, **85**, 177
Jeffreys, W.H. 1981, *AJ*, **86**, 149
Kirkpatrick, S., Gelatt, C.D., Jr., & Vecchi M.P. 1983, *Science*, **220**, 671
Levenberg, K 1944, *Quart. App. Math*, **2**, 164
Marquardt, D.W., 1963, *J. Soc. Ind. Appl. Math*, 11, 431
Morbey C.L. 1975, *PASP*, **87**, 689
Morbey, C.L. & Griffin R.F., 1987, *ApJ*, **317**, 343
Press, W.H., Flannery, B.P., Teukolsky S.A., & Vetterling, W.T. 1986, *Numerical–Recipes: The Art of Scientific Computing*, (Cambridge, Cambridge University Press)

8.   DISCUSSION

HARRINGTON: Even in just the visual orbit computation problem, where we have both visual and photographic data, we have the same mixed–data problem of weighting, correlations, and so forth. These considerations are with us everywhere.