A METHOD FOR COMPARING DISCRETE KINEMATIC DATA AND N-BODY SIMULATIONS

PRASENIT SAHA

Mount Stromlo and Siding Spring Observatories, Australian National University, Weston Creek, ACT 2611, Australia

Received 1997 September 3; revised 1998 October 29

ABSTRACT

This paper describes a method for quantitatively comparing an N-body model with a sample of discrete kinematic data. The comparison has two stages: (1) finding the optimum scaling and orientation of the model relative to the data, and (2) calculating a goodness of fit and hence assessing the plausibility of the model in view of the data. The method is derived from considering both the data and model as samples from some underlying binned distribution function and applying probability theory arguments. As an example, I consider a published N-body model for the Galactic bulge and disk, and fictitious \((l, b, v)\) measurements, and recover (with error estimates) the spatial and velocity scales of the model and the orientation of the bar. The fictitious data are actually derived from the model by assuming the mass scale and the solar position, but their size and extent mimic a recent survey of OH/IR stars. The results indicate that the mass of the bulge and our viewing angle of the bar are usefully estimable from current surveys.

Key words: galaxies: kinematics and dynamics — methods: analytical — methods: numerical

1. INTRODUCTION

Kinematic surveys of a population of discrete objects are an increasingly important kind of data in galactic astronomy. The objects may be stars in globular clusters or elsewhere in the Galaxy (e.g., Meylan & Mayor 1986), emission-line objects in the Galaxy or other galaxies (e.g., Ciardullo et al. 1993; Hui 1993; Arnaboldi et al. 1994; Tremblay, Merritt, & Williams 1995; Beaulieu 1996; Sevenster et al. 1997a, 1997b), or galaxies in a cluster (e.g., Colless & Dunn 1996). Such surveys usually measure sky positions and line-of-sight velocities, but for some systems, proper motions are also available (e.g., Spaenhauer et al. 1992).

One would like to be able to throw these data at some dynamical analysis machine and reap all the dynamical results implicit in the data, but there is no such machine. Some progress toward this goal has been made, notably by D. Merritt and collaborators (see Merritt 1993; Merritt & Tremblay 1993; Merritt & Gebhardt 1994; and especially Merritt 1996). These papers develop methods for reconstructing mass profiles (including dark matter) from kinematic observations, in a model-independent way. But at present, they extend only to axisymmetric systems viewed in the equatorial plane. So for triaxial systems, and certainly for nonequilibrium systems like clusters of galaxies, it is basically N-body simulations that have to be confronted with data. How can we best do this quantitatively?

Generally speaking, there are three questions to which one would like answers when comparing N-body models with observations:

1. How should a model be scaled and oriented to best fit the data?
2. Could the data at hand have plausibly come from a particular model’s distribution, or do the data rule out the model?
3. If there are several plausible models, which one do the data favor?

All three questions are answerable if we can calculate the likelihood function, which is the probability of having gathered the actual data under a particular model. Suppose for definiteness that the data consist of measurements of sky position \((l, b)\) and line-of-sight velocity \(v\), with negligible errors. Let us also assume for now that the simulation is so fine-grained that it effectively gives us a distribution function \(f\). One usually thinks of \(f\) as a function of phase-space variables, but we can change variables to express it as \(f(l, b, v, \eta)\), where \(\eta\) stands for three unmeasured numbers (e.g., distance and proper motion). Then the probability of drawing values \((l_k, b_k, v_k)\) from \(f\) is

\[
\text{prob} \left( l_k, b_k, v_k \mid f \right) = \int f(l_k, b_k, v_k, \eta)d\eta.
\]

Assuming the data on different objects are independent, we have for the likelihood

\[
\text{prob} \left( \text{data} \mid f \right) = \prod_k \left[ \int f(l_k, b_k, v_k, \eta)d\eta \right].
\]

Since \(f(l, b, v, \eta)\) will depend on the scalings and orientation adopted, we can fit for these parameters—the peak of prob \((\text{data} \mid f)\) in the relevant parameter space estimates the parameters, and the broadness of that peak gives uncertainties. To answer question 2, we can test whether the value of prob \((\text{data} \mid f)\) is typical of random data sets drawn for that \(f\); if prob \((\text{data} \mid f)\) is anomalously low, we can infer that \(f\) is inconsistent with the data. Question 3 can be answered by comparing prob \((\text{data} \mid f)\) for the various models available; there is an extra complication though, in that we must marginalize over the parameters for each model—see Sivia (1996) for a discussion of this point. I will not address model comparison in this paper.

The contribution of this paper is to derive and test a practical approximation to the “in principle” procedure described above. We need an approximation because particle simulations do not give us \(f\) directly; we need to smooth somehow. Smoothings in general introduce biases, so we have to monitor for biases and correct for them if necessary. But bearing that caution in mind, the smoothing that I propose to use is the simple-minded one of just binning in \((l, b, v)\), i.e., assuming that \(f\) is constant within boxes in

---

1 Current address: Department of Physics (Astrophysics), University of Oxford, Keble Road, Oxford OX1 3RH, England, UK; saha@physics.ox.ac.uk.
(l, b, v)-space. Let us say that for some choice of scaling and orientation parameters, the lth bin has \( m_l \) model points and \( s_l \) data object points; also let \( M = \sum_i m_i \) and \( S = \sum_i s_i \). This immediately suggests minimizing \( \chi^2 \) to obtain a best fit, but that is a bad idea. Minimizing \( \chi^2 \) implicitly assumes that the \( s_l \) follow a Gaussian distribution, the mean and variance in this case being both equal to \( m_l S/M \). This is fine if \( s_l \gg 1 \), but \( S \) having values typically in the dozens to hundreds, we do not have the luxury. Moreover, for bin sizes of interest, even the \( m_l \) may not always be large enough for shot noise to be negligible. The solution is to view both the sets \( m_l \) and \( s_l \) as samples drawn from some underlying \( \text{Gaussian data.} \)

The derive from data, and for a given data set and \( m \) and \( s \), the likelihood then takes the form

\[
\text{prob (data | model)} \propto W = \prod_{i=1}^{B} \left( \frac{(m_i + s_i)!}{m_i!s_i!} \right) .
\]

This formula is derived in the Appendix, but note two intuitively desirable properties of \( W \): (1) the \( (m_i + s_i)! \) factor favors large \( m_i \) coinciding with large \( s_i \), but the denominator discourages extremes like \( m_i = M \) at the bin with highest \( s_i \) and zero elsewhere; (2) if some outlier observation lands in a bin with no model points (i.e., \( m_i = 0 \), \( s_i = 1 \)), that bin contributes unity to the product—in this sense \( W \) is robust against outliers.

Although equation [3] is symmetric in \( m_i \) and \( s_i \), operationally these two sets of numbers will play quite different roles. The \( s_i \) derive from data, and for a given data set and binning, they are fixed. The \( m_i \), on the other hand, depend on the scaling and orientation parameters and will vary as those parameters are adjusted to \( W \).

To explain the details of the use of \( W \), it is probably best to work through an example; below we work through the problem of scaling and orienting \( N \)-body models of the Milky Way bulge and inner disk from \( (l, b, v) \) measurements.

As it happened, it was this problem that led to the present work, but the bulge is a good example to illustrate anyway, for two reasons. First, it is a triaxial system with the interstellar medium (ISM) as the viewing angle of the bar, and \( \varphi \) is the viewing angle of the bar, and our radial velocity is assumed to equal 0 or is corrected for. In the convention implied by equations [4] and [5], a value of \( \varphi \) between 0° and 90° signifies that the nearer side of the bar is at positive \( l \) and (because the model has positive rotation) positive \( v \). The real Galactic bar is believed to be in such an orientation.

The asymmetry between the spatial and velocity components of equation [4] may seem odd—why do we not have

\[
x' = r_{\text{scale}} (x \cos \varphi - y \sin \varphi) , \\
y' = r_{\text{scale}} (x \sin \varphi + y \cos \varphi) + R_0 , \\
z' = r_{\text{scale}} z 
\]

We would need an extra parameter \( r_{\text{scale}} \) if we were considering proper-motion data (available for Baade’s window stars in Spahnhauser et al. 1992). But from equation [5] the observables all depend only on the ratio \( R_0 / r_{\text{scale}} \), so in equation [4] we drop \( r_{\text{scale}} \). Then \( R_0 \) in effect becomes a surrogate for the spatial scale: if our true Galacticentric
distance is 8.5 kpc, then
\[ r_{\text{scale}} = \left( \frac{8.5}{R_0} \right) \left( \frac{\text{kpc}}{\text{model unit}} \right). \]

With proper-motion data, in principle \( r_{\text{scale}} \) and \( R_0 \) could both be determined, thus providing the actual Galactocentric distance; but with only (\( l, b, v \)) that distance must be supplied separately to obtain \( r_{\text{scale}} \). Once we have \( r_{\text{scale}}, \) we can find the mass, because the scale for \( G \times \text{mass} \) is \( r_{\text{scale}} \times v^2 \) (since \( GM \) has dimensions of \( L^3 T^{-2} \)).

Consider now a survey of (\( l, b, v \)) measurements, which we would like to compare with the simulation and infer \( R_0, \phi, r_{\text{scale}}, \) and \( v_0 \) to the extent possible. The first step is to choose the bins in (\( l, b, v \)) for comparison—for more on choosing bins, see below, but for the moment suppose we have chosen our \( B \) bins. This sets the \( s_i \). The \( m_i \) will depend on what scaling and orientation parameters we have chosen; for any choice we can put the model particles through the transformations in equations [4] and [5], bin them, and randomly pick \( M \) out of all the particles that fall into our \( B \) bins, thus obtaining \( m_i \). Then we calculate \( W \), which clearly depends on the parameters.\(^2\) Clearly, our strategy for estimating the parameters will be to vary them so as to maximize \( W \). Obtaining error bars and testing the model are more involved and are discussed in detail in § 3. They will involve simulating data sets from the model. Generating a simulated data set \( s_i \) from the model is like generating \( m_i \), except that we choose \( S \) particles rather than \( M \). Sometimes we will be calculating \( W \) for two sets of occupancies \( s_i \) and \( m_i \), both of which come from the model, but using different parameter values.\(^3\)

How do we choose the bins? Because of the assumptions that go into the derivation of \( W \), it is best to avoid unequally sized bins. But there is no need to have bins in unsurveyed regions, so bins need not be contiguous. The bin size requires some thought. I cannot suggest any definite prescription for the bin size, but there are two guidelines. First, \( B \) should be several times smaller than \( M \), so that \( m_i \) can be large enough to actually carry some information about the distribution function; \( B \approx M/5 \) seems servicable. There is no problem with \( B \gg S \); after all, the continuous limit is \( M \gg B \gg S \). Second, the binning should not be so coarse that it misses important features in the distribution function. Too coarse a binning can lead to strange biases, as the following suggests: The scale height of the bulge (as seen from the solar system) is about 2:2: suppose the bins were 5° in \( b \). When fed data binned thus, any model-fitting procedure is likely to respond by fitting a model with an increased scale in \( b \). An easy way to do this is to increase \( R_0 \)—but then the fit would have to compensate for the scale in \( l \), which it might do by reducing \( \phi \) to make the bar more

\(^2\) The number of model particles that fall into our \( B \) bins will depend on the parameters. Particles may fall outside the survey region, where we might not have bins. But \( M \) must be maintained the same for all parameter values, i.e., we must always choose a subset of size \( M \) of those model particles that do fall into our \( B \) bins. Otherwise eq. [3] for \( W \) becomes invalid (see Appendix).

\(^3\) If we are going to compare mock data generated from a model with that model, it is important to then remove from the model those points that composed the mock data. A model particle should never contribute to both \( s_i \) and \( m_i \) at the same time. The reason is that the data are not supposed to correspond exactly to any model particles, only to have come from the same distribution function.

3. USE OF THE \( W \)-FUNCTION

It is straightforward to incorporate the likelihood \( W \) in standard Monte Carlo procedures for parameter estimation and model testing, and the following describes how this can be done. The approach here is not the only possible one, and Bayesian purists would reject it entirely; but it seems computationally to be the most tractable.

It is helpful to consider two functions, \( D \) and \( \Omega \). \( D(\omega) \) is a function that generates a data set from the model using parameter values \( \omega \). \( D \) is probabilistic, so there can be many possible data sets \( D_1, D_2, \ldots, (\omega) \) from the same model and parameters. If the model is correct and the true values are \( \omega_{\text{true}} \), then the observed data can be thought of as one realization of \( D \):

\[ D_{\text{obs}} = D(\omega_{\text{true}}). \]

\( W \) depends on both the data and the parameters:

\[ W = W(D, \omega') \text{ or } W[D(\omega), \omega']. \]

In general, \( \omega' \neq \omega \); \( \omega \) leads to the data and hence to \( s_i \), but \( m_i \) are derived by applying a possibly different value \( \omega' \) to the model. We now define the function \( \Omega \) thus:

\[ \Omega(D) = \omega': W(D, \omega') \text{ is maximum}. \]

To calculate \( \Omega(D) \), we do not need to know what \( \omega \) value gave \( D \); but \( \Omega(D) \) is an estimator for that unknown value (in fact, a maximum likelihood estimator, because \( W \) is a likelihood).\(^4\)

To estimate parameters, we calculate \( \omega_{\text{est}} = \Omega(D_{\text{obs}}) \). For error bars on \( \omega_{\text{est}} \) we want the scatter in \( \Omega[D(\omega_{\text{true}})] \). But since in real applications we will not know \( \omega_{\text{true}} \), we can take instead the scatter in \( \Omega[D(\omega_{\text{est}})] \); from this we can read off desired confidence limits. This is a standard Monte Carlo error estimation—see Figures 15.6.1 and 15.6.2 in Press et al. (1992).

Testing the model is slightly more complicated. We need some statistic that measures the goodness of a parameter fit, but the well-known ones do not help us: \( \chi^2 \) is inappropriate for the reasons given in § 1, and the Kolmogorov-Smirnov (K-S) statistic and its relatives are inapplicable, because the data are not one-dimensional. However, there is an obvious choice: \( W \) itself. To apply it, we compare \( W[D_{\text{obs}}, \omega_{\text{est}}] \) with the distribution of \( W[D(\omega_{\text{true}}), \omega_{\text{est}}] \). If \( W[D_{\text{obs}}, \omega_{\text{est}}] \) lies in the lowest percentile of the distribution of \( W[D(\omega_{\text{true}}), \omega_{\text{est}}] \), then the model is rejected at 99% significance, and so on. For \( \chi^2 \) and also for K-S test and its relatives, the distribution corresponding to \( W[D(\omega_{\text{true}}), \omega_{\text{est}}] \) is model independent. In our case, we will need to calculate the distribution; but again, we will not know the true \( \omega_{\text{true}} \), so we will have to substitute the distribution of \( W[D(\omega_{\text{est}}), \omega_{\text{est}}] \).

Note that if the goodness-of-fit test leads to a rejection of the model, parameter estimates from that model must be

\(^4\) If we had \( \langle \Omega[D(\omega)] \rangle = \omega \) (the average being over an ensemble of \( D \) with \( \omega \) fixed), then \( \Omega \) would be an unbiased estimator. As suggested in § 1, in practice, \( \Omega \) will have some bias, because the binning process introduces bias. But that is not a problem, provided we can test for bias and correct for it where required.
from current surveys, and how well. I used 21 part of the survey of bulge OH/IR stars by Sevenster et al. (1998, IG 320 km s \(^{-1}\)).

4. SIMULATIONS WITH SELLWOOD’S MODEL

In this section, I present Monte Carlo simulations to gain some idea of which bulge parameters can be constrained from current surveys, and how well.

Consider Sellwood’s model with \( \omega_{\text{true}} \) being \( R_0 = 6 \) model units, \( \varphi = 30^\circ \), \( v_0 = 220 \) km s \(^{-1}\), and \( v_{\text{scale}} = 300 \) km s \(^{-1}\) per model unit. All these are plausible values, and using them I computed \( \Omega[D(\omega_{\text{true}})] \) for 32 mock surveys, each having 300 objects in the range \( |l| < 10^\circ \), \( |b| < 3^\circ \), \( |v| < 320 \) km s \(^{-1}\). The size and extent mimic the symmetric part of the survey of bulge OH/IR stars by Sevenster et al. (1997a, 1997b; the survey region need not be symmetric—see below). I used \( 21 \times 13 \times 16 \) bins in \( (l, b, v) \). Figure 2 shows the results: the ranges of the axes in \( R_0 \), \( v_0 \), and \( \varphi \) are the ranges I searched; for \( v_{\text{scale}} \) I searched the range 0–500.

Figure 3 shows a sort of direct visual data-model comparison for the first of the mock surveys. Such plots are useful in checking for programming goofs, but otherwise there is not much information one can extract from them.

Figure 4 illustrated that the maximum of \( W \) of the first mock survey is typical of the values one should expect from this model, as expected since the mock survey came from the model. With real data, such a plot would test whether the data could plausibly have come from the model being studied.

From the results in Figure 2, the medians and 68\% range in \( \Omega[D(\omega_{\text{true}})] \) are \( \varphi = 29^{+8}_{-7} \) deg, \( v_{\text{scale}} = 290^{+20}_{-13} \). These numbers indicate the sort of bias and error bars we can expect from a survey of this size and extent. We see that \( v_{\text{scale}} \) can be estimated to \( \sim 10\% \) and \( \varphi \) to \( \sim 10^\circ \) with no need to correct for bias. On the other hand, we obtain no useful information on \( v_0 \) and \( R_0 \). That \( v_0 \) is not constrained by such data is not surprising, since it is almost perpendicular to what is measured. But the inability to infer \( R_0 \) is puzzling, especially considering the impressively tight constraint, on \( \varphi \). Evidently, we must use integrated light to estimate \( R_0 \).

\[ \text{Fig. 3.—Open circles show (l, b) and (l, v) for a mock survey } D_s(\omega_{\text{true}}). \text{ The filled circles show (l, b) and (l, v) for a mock survey } D_1(\omega_{\text{true}}), \text{ where } \omega_{\text{true}} = \Omega[D_1(\omega_{\text{true}})]. \text{ A real survey would use } D_{\text{obs}} \text{ instead of } D_1(\omega_{\text{true}}). \]

\[ \text{Fig. 4.—Vertical dashed line shows the value of } \ln W[D_s(\omega_{\text{true}}), \omega_{\text{true}}], \text{ where } \omega_{\text{true}} = \Omega[D_s(\omega_{\text{true}})] \text{ which is marked in Fig. 2. With a real survey } D_s(\omega_{\text{true}}) \text{ would be replaced by } D_{\text{obs}}. \text{ The rising curve is the cumulative distribution of } \ln W[D_1, \ldots, D_{100}(\omega_{\text{true}}, \omega_{\text{true}})]. \text{ As expected, } D_s(\omega_{\text{true}}) \text{ appears as typical of the distribution } D_s(\omega_{\text{true}}). \]
It is interesting to observe what happens as the surveys get larger. Figure 5 shows $\Omega[D(\omega_{\text{true}})]$ for mock surveys when the size is extended to 500 and the $l$ range is extended to $-45^\circ < l < 10^\circ$, which mimics the full size and extent of Sevenster et al.'s (1997a, 1997b) OH/IR survey. We now find approximate medians and 68% ranges of $\varphi = 29^{\circ} \pm 8^\circ$ deg and $v_{\text{scale}} = 290^{\circ} \pm 23^\circ$, but the outliers are noticeably less distant. We also note a bias toward low estimates for $v_{\text{scale}}$ and high estimates for $R_0$; if the survey size is increased further, the scatter in $\Omega[D(\omega_{\text{true}})]$ reduces further, and the biases in $v_{\text{scale}}$ and $R_0$ become correspondingly more noticeable.

To summarize the results, these survey simulations indicate that current surveys can constrain the viewing angle of bulge simulations to less than $10^\circ$ and the velocity scale to less than 10% (at 68% confidence). The spatial scales will need to be set independently, using integrated light. Kalnajs (1997, private communication) obtains $\sim 10\%$ or better constraints on $R_0$ by comparing $N$-body models and integrated light from COBE. Combining with less than 10% uncertainties on $v_{\text{scale}}$, it appears that $N$-body models could be scaled in mass to $\sim 25\%$. The resulting predictions for microlensing optical depths would easily be tight enough for interesting confrontations with bulge microlensing observations.

I am grateful to Ken Freeman for posing this problem and to Sylvie Beaulieu and Maartje Sevenster for teaching me about bulge observations. Thanks also to Agris Kalnajs and Maartje Sevenster for the appropriate mixture of enthusiasm and skepticism about $W$.

### APPENDIX

This appendix derives the likelihood formula (eq. [3]) from probability theory arguments. Continuing the notation of § 1, we suppose that there are $B$ bins in all, and that in the $i$th bin $f_i = f_i$ and both $m_i$ and $s_i$ are drawn from $f_i$.

The joint probability for the bin occupancies is the product of two multinomial distributions:

$$\text{prob} \left( s_i, m_i \mid f_i \right) = M!S! \prod_{i=1}^{B} \frac{f_i^{m_i + s_i}}{m_i!s_i!}. \quad (A1)$$

Equation (A1) could be used to estimate the $f_i$, with uncertainties. But as we have no particular interest in the $f_i$ as such, we marginalize them out in the usual way, which is to integrate over all allowed values of the $f_i$—that means all combinations of values of the $f_i$ between 0 and 1, subject to $\sum f_i = 1$. Distributions of $f_i$ that most resemble the $m_i$ and $s_i$ will, according to equation (A1), contribute most to the integral. Marginalization is just an application of the additive rule for probabilities. Using the identity

$$\left( \prod_{i=1}^{B} \int f_i^m \, df_i \right) \delta \left( \sum f_i - 1 \right) = \frac{1}{(N + B - 1)!} \prod_{i=1}^{B} n_i!, \quad (A2)$$

we obtain

$$\text{prob} \left( s_i, m_i \right) = \frac{M!S!(B - 1)!}{(M + S + B - 1)!} \prod_{i=1}^{B} \frac{(m_i + s_i)!}{m_i!s_i!}. \quad (A3)$$

Setting $S = 0$ in equation (A3) gives $\text{prob} \left( m_i \right)$, and hence

$$\text{prob} \left( s_i \mid m_i \right) = \frac{S!(M + B - 1)!}{(M + S + B - 1)!} \prod_{i=1}^{B} \frac{(m_i + s_i)!}{m_i!s_i!}. \quad (A4)$$

If $B = 1$, then the probabilities in equations (A3) and (A4) become unity, as they should. For the sort of applications of interest in this paper, $M, S, B$ and $B$ are fixed, so we can ignore the normalization in equation (A4) and worry only about $W$ as defined in equation (3).
If $M$ is large enough compared with $S$ that $m_i + s_i \approx m_i$ and equation (A4) simplifies to
\[
\text{prob } (s_i | m_i) \propto \prod_i \frac{m_i^{s_i}}{s_i!},
\] (A5)
which amounts to saying that $f_i = m_i / M$ because the shot noise in the $m_i$ is negligible. If $M$ is large enough that we can make the bins so small that each $s_i = 0$ or 1, but still all $m_i \gg 1$, then equation (A5) simplifies further, to
\[
\text{prob } (s_i | m_i) \propto \prod_{j: s_j = 1} m_j,
\] (A6)
which amounts to equation (2) for the continuous case. Thus equation (A4) has the expected large-$m_i$ limits.

REFERENCES

Arnaboldi, M., Freeman, K. C., Hui, X., Capaccioli, M., & Ford, H. 1994, Messenger, 76, 40
Beaulieu, S. F. 1996, Ph.D. thesis, Australian Natl. Univ.
Charbonneau, P. 1995, ApJS, 101, 309
Ciardullo, R., Jacoby, G. H., & Dejonghe, H. B. 1993, ApJ, 414, 454
Colless, M., & Dunn, A. M. 1996, ApJ, 458, 435
Fux, R. 1997, A&A, 327, 983
Gerhard, O. E. 1996, in Proc. IAU Symp. 169, Unsolved Problems of the Milky Way, ed. L. Blitz & P. Teuben (Dordrecht: Reidel), 79
Hui, X. 1993, PASP, 105, 1011
Merritt, D. 1993, ApJ, 413, 79
———. 1996, AJ, 112, 1085
Merritt, D., & Gebhardt, K. 1994, in Proc. of 29th Rencontre de Moriond, Clusters of Galaxies, ed. F. Durret, A. Mazure, & J. Trần Thanh Vân (Gif-sur-Yvette: Ed. Frontières), 11
Merritt, D., & Tremblay, B. 1993, AJ, 106, 2229

Meylan, G. & Mayor, M. 1986, A&A, 166, 122
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in FORTRAN (2d ed; Cambridge: Cambridge Univ. Press)
Sellwood, J. A. 1993, in AIP Conf. Proc. 278, Back to the Galaxy, ed. S. S. Holt & F. Verter (New York: AIP)
Sevenster, M. N., Chapman, J. M., Habing, H. J., Killeen, N. E. B., & Lindqvist, M. 1997a, A&AS, 122, 79
———. 1997b, A&AS, 124, 509
Sivia, D. S. 1996, Data Analysis: A Bayesian Tutorial (Oxford: Clarendon)
Spaenhauer, A., Jones, B. F., & Whitford, A. E. 1992, AJ, 103, 297
Te Lintel Hekkert, P., Caswell, J. L., Habing, H. J., Haynes, R. F., & Norris, R. P. 1991, A&AS, 90, 327
Tremblay, B., Merritt, D., & Williams, T. B. 1995, ApJ, 443, L5
Zhao, H. S. 1996, MNRAS, 283, 149