SPHERICAL GALAXIES: METHODS AND MODELS

Douglas O. Richstone
Department of Astronomy
University of Michigan

ABSTRACT. Over the last 5 years, considerable progress has been made in our ability to construct self-gravitating stellar equilibria. One of these new methods is essentially a variant of Eddington's (1916) method. Two other key approaches are logical extensions of Schwarzschild's Linear Programming method, and can be applied to nonspherical models as well. These methods are reviewed below.

The application of these methods to galaxies has yielded a few very interesting results within the last year or two. The methods described below unambiguously establish M/L's for M87 and M32 within about 30 arc seconds. They strongly support Tonry's contention that the nucleus of M32 contains a large invisible mass, possibly a $10^7 M_\odot$ black hole. They also suggest that observational recovery of the projected velociy distribution might permit the observer to distinguish between a massive halo and an increasingly tangential velocity distribution function.

1. METHODS

We begin with Jeans's theorem, which states that the phase space distribution function $f(\mathbf{r},\mathbf{v})$ is a function only of the isolating integrals. For truly spherical symmetry these are the energy (E) and the square angular momentum (J^2). Any dependence on the individual components of J other than through J^2 would create a prefered dynamical direction. In a sense, Jeans's theorem solves the problem of constructing spherical systems: we simply choose a form of the phase space distributioin function $f(E, J)$, integrate over all velocities to find the density ρ, and then solve the Poisson Equation for the potential Φ, which is contained implicitly in f and ρ.

This approach in fact describes the first class of methods — which we term King type methods (see Michie and Bodenheimer 1963 and King 1965). It has been recently applied to the dynamics of clusters of galaxies by Kent and Gunn (1982). The method requires a reasonably simple choice of the functional form of f, so that the integral for the density can be performed in terms of the potential. Its major limitation is that there is no guarantee that the assumed distribution function approximates the system under study.

Moreover, under normal circumstances we have a density profile $\rho(r)$ and wish to recover the distribution function which gives rise to that spatial density. This problem is underdetermined, since we are deriving a function of two variables (E and J) from a function of one variable. Eddington (1916) provided one possible solution

T. de Zeeuw (ed.), Structure and Dynamics of Elliptical Galaxies, 261–269.

https://doi.org/10.1017/S0074180900185237 Published online by Cambridge University Press

by showing that if $f = f(E)$ it can be recovered from $\rho(r)$. A distribution function which is only a function of energy is isotropic and contains no free parameters. Merritt (1985) has shown that the distribution function can also be recovered from the density if it is a function of $Q_{\pm} = E \pm J^2/r_a^2$. This choice of functions has two advantages over Eddington's: it permits the modelling of velocity anisotropy and it contains an adjustable parameter. For many choices of $\rho(r)$, it will yield physically reasonable (nonnegative) distribution functions. Even so, its virtue is also its fault: the explicit restriction to a particular combination of the two integrals sharply restricts the solutions which can be found. It also will happily produce distribution functions which are negative in some regions of phase space, although it is immediately obvious when this occurs.

A third useful approach to this problem was invented by Binney and Mamon (1982, see also Tonry 1983), in their study of M87. Binney and Mamon used the density profile, observed velocity dispersion profile, and the equation of stellar hydrodynamics to recover the internal radial and tangential velocity dispersions. Unfortunately, this method does not work well when the observed dispersion is available over a small range in radius. It also has the important limitation that the dispersions constructed by this method need not correspond to a nonnegative distribution function, and it may in fact not be easy to discover that. For example, it is possible to use this technique to construct a de Vaucouleurs law model with purely radial orbits by setting the tangential dispersions everywhere equal to zero; such a system does not have a nonnegative distribution function everywhere (see Richstone and Tremaine 1984).

A fourth kind of approach to this problem was pioneered by Schwarzschild (1979). He noted that each time averaged orbit in any specific potential is a solution of the collisionless Boltzmann equation, and that these orbits can be summed to find a solution of the Poisson equation for that potential. He chose to use linear programming to perform the sums, since it guarantees a set of orbit occupation numbers with nonnegative weights. The formal validity of this approach has been demonstrated by Vandervoort (1984). In the spherical case, this approach can be described as writing the distribution function as a sum of products of delta functions of the isolating integrals: $f(E, J^2) = \sum_i [\delta(E - E_i)\delta(J^2 - J_i^2)]$. It is immediately clear that one can in fact express f as any sum of functions of isolating integrals. In this regard Merritt's functions may be particularly useful, since any weighted sum of his functions (say, with different r_a's) will remain a valid solution for that particular $\rho(r)$.

Although linear programming has been used extensively for modelling spherical galaxies (Richstone and Tremaine 1984, 1985), there are at three other methods which can be used to combine orbits (or families of phase space distribution functions) to produce models. Each of these methods must solve (or approximate) a set of equations of the form

$$M_i = \sum_i m_{ij} w_j, \qquad (1)$$

$$w_j \geq 0,$$

which states that the sum of the products of the mass distributions of the orbits (m_{ij}) with their occupation numbers (w_j) is equal to the mass distribution (M_i) of the galaxy under consideration. One such method is advocated by Pfenniger (1984) and described in detail in Lawson and Hanson (1974). It uses a pivoting method (like linear programming) to minimize the sum of squared residuals from eqn (1) above.

A second method is the application of Lucy's (1974) method due to Newton and Binney (1984). Lucy's method solves equation 1 by starting with some guess for w_j with no zero or negative components, and iterating as follows:

$$Q_{ij} = \frac{m_{ij} w_j^{(g)}}{\sum_i m_{ij} w_j^{(g)}}, \tag{2}$$

$$w_j^{(g+1)} = \sum_i M_i Q_{ij}. \tag{3}$$

Lucy's method is motivated by Bayes's theorem on conditional probabilities. This method turns out to work extremely well for good first guesses. Its great virtue relative to linear programming is that it tends to produce smooth distribution functions. It is also fairly easy to code.

Another method which has been used to construct galaxy models is a maximum entropy method using Lagrange multipliers to stay on the constraints (Richstone and Tremaine 1986). Any entropy of the form $\sum_j S_j(w_j)$ can be maximized subject to those constraints by solving

$$S_j' - \sum_i \lambda_i m_{ij} = 0. \tag{4}$$

subject to equation 1, where $S_j' = \partial S_j / \partial w_j$. This can be accomplished by guessing w_j , and expanding eq (4) as a Taylor series to get

$$\Delta w_j = \frac{\sum_i \lambda_i m_{ij} - S_j'}{S_j''}, \tag{5}$$

for Δw_j and multiplying by m_{kj} and summing to get

$$\sum_i \lambda_i \Big(\sum_j \frac{m_{ij} m_{kj}}{S_j''} \Big) = \sum_j \frac{S_j'}{S_j''} m_{kj} + \Delta M_k, \tag{6}$$

for λ_i. Eqn 6 is solved first for the λ_i and eqn 5 is then solved for the Δw_j . This method converges very rapidly, but requires the solution of a set of linear equations (for the λ_j) at each step.

Statler (unpublished) has shown that the Binney-Newton-Lucy method leads to a variety of different solutions in the solution space. We have compared the operation of this method and the maximum entropy method for a toy problem and found that the maximum entropy method usually finds a solution in about 1/5 the number of iterations of the BNL method, but then spends more time finding the particular solution with maximum entropy. The virtue of the maximum entropy method is that it produces a solution to the problem which is well defined in terms of some principle.

It does not seem appropriate to regard the classical entropy $f\ln(f)$ as having any particular significance for galaxies. They are, after all, only violently relaxed (see Tremaine, Henon, and Lynden-Bell 1986). Their further diffusion in phase space occurs on a timescale much longer than the age of the Universe. In this view, the entropy serves as a device to avoid negative occupation numbers and produce a smooth distribution function. For this purpose, other choices of an 'entropy' function would suffice, and we have explored other choices. All models displayed below were constructed using the classical definition.

2. MODELS

Various authors have employed these methods to construct spherical galaxy models during the last few years. Here I first want to display two maximum entropy models (using the classical entropy) with mass distributions of a Plummer Model and of an $r^{1/4}$ law. Both models were rescaled in mass to have the same observed velocity dispersion. Note that in each case the dispersion profile slightly favors tangential anisotropy at large radii. In these cases, the maximum entropy method described above converges to a solution in less than 4 iterations, and finds a maximum to reasonable precision in 10 iterations.

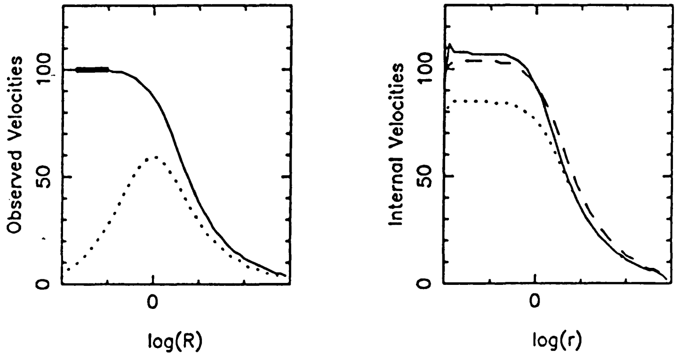

Figure 1. – Dynamical properties of a maximum entropy Plummer model. Left box shows the observed dispersion for a nonrotating model (solid line), and the maximum rotation rate as described in the text (dotted line) as a function of projected radius. The right box show the σ_r, σ_t for a nonrotating model and the maximum rotation rate as a function of radius.

One important use of these methods is to improve our understanding of galactic nuclei. Since the most popular theory of quasars uses supermassive black holes for energy production, the demonstration that they are present in galactic nuclei would provide strong support for that view. Starting with Young et. al. (1978) and Sargent et. al. (1978) various investigators have addressed these issues. Tremaine and I (1985) have recently reviewed the history of work on M87. For that particular galaxy, only one additional footnote seems appropriate here. Not only (as Duncan and Wheeler(1980) and Binney and Mamon pointed out) is it possible to construct constant M/L models for the galaxy. It is even (especially if Dressler's(1980) results for the dispersion near the center are correct) easy to make a variety of constant M/L models consistent with the observations, including models with only mildly anisotropic velocity dispersions. Such a model is displayed below.

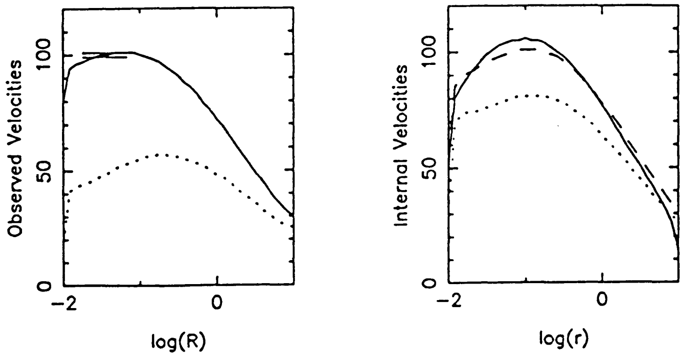

Figure 2. – Same as Figure 1 for a deVaucouleurs law mass distribution.

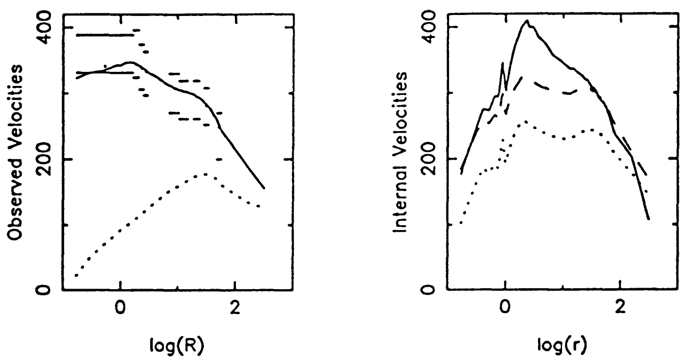

Figure 3. – A dynamical model for M87. The double horizontal lines in the left box are the 1σ upper and lower error bars from observations. Otherwise the quantities plotted are the same as in Figures 1 and 2. The units are km/sec and arcseconds.

The situation is much more interesting in the case of M32, which has been studied by Tonry (1984) and Dressler (1984) and now by Kormendy (this meeting). M32 displays a sharp jump in velocity dispersion, which may be unresolved rapid rotation. Tonry remarked that the rotation curve of M32, which reaches about 40 km/sec at about 3 arcseconds (12 pc) from the center, must be lowered by seeing. Seeing has a particularly significant effect on rotation curves, since it carries light from the wrong side of the minor axis into the slit. Tonry modeled the rotation curve with various profiles, concluding that the true rotation at about 3 arcseconds was about 70 km/sec and that it must rise still further at smaller radii. He stated that this established a higher M/L inside 2 arcseconds than outside 2 arcseconds.

I have attempted to verify this statement by using the maximum entropy method to construct M32 models consistent with the observed velocity dispersions with the maximum amount of rotation at the center consistent with a spherical model. The maximum rotation rate for a given orbit at a given point can be shown to be $2/\pi(J_t^2/r^2)^{1/2}$. The M32 model with maximum rotation near 1 arcsecond is shown below. Note that the projected rotation velocity near 1-3 arcseconds is only 60 km/sec . I felt that this discrepancy of 60 vs 70 km/sec was too small to justify Tonry's claims for this object, but now that Kormendy has demonstrated that the rotation velocity of the galaxy must continue to rise inside 2 arcseconds it is clear that the M/L must rise above the 2.4 ± 0.1 (red) that characterizes larger radii in this galaxy. This makes M32 a prime candidate for a black hole in the $10^7 M_\odot$ range.

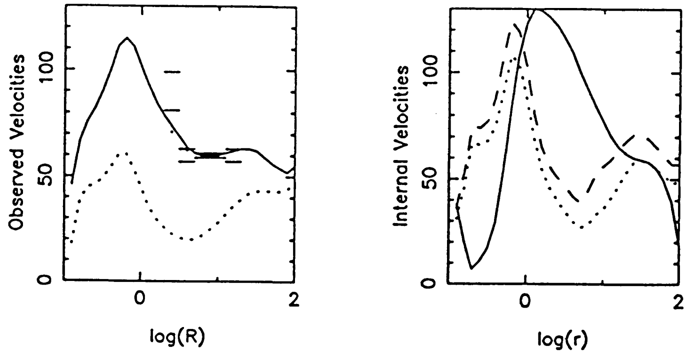

Figure 4. – Same as Figure 3 for M32.

In retrospect, M31 and M32 should have been studied all along, since the maximum radius at which such an object can make its influence felt is defined by

$$r = \alpha G M_h / v^2, \tag{7}$$

where M_h is the mass of the black hole, v characterizes the velocities of the stars near the center of the galaxy, and α is of order unity. So, in order to 'see' the black hole in a galaxy, it must have a minimum mass of order

$$M_h = \frac{\sigma^2 D\theta}{\alpha G}, \tag{8}$$

where D is the distance to the galaxy, θ is the observational spatial resolution and σ is the system's velocity dispersion. M32, with its dispersion of 60 km/sec and its distance of 700 kpc, presents a minimum detectable black hole mass near $10^6 M_\odot$, while M87, at 15Mpc and $\sigma = 300$ km/sec has a minimum mass near $1.5 \times 10^9 M_\odot$, (for $\theta = 1$ arcsec) so a $10^8 M_\odot$ object in M87 will never be dynamically detectable from the ground.

An interesting sidelight of the careful work on M87 and M32 is that the systemic M/L ratios for these systems are now quite well constrained, from about 2 arcseconds to about 30 arcseconds, assuming that the mass follows the light over

https://doi.org/10.1017/S0074180900185237 Published online by Cambridge University Press

that range in radius. In that case, M/L for M87 is 10 ± 2 (in B), while for M32 it is 2.4 ± 0.1 in Gunn red. This real difference in M/L is not an artifact of using different bandpasses. It cannot be wriggled out of via anisotropy. It must reflect a real difference in the stellar populations of these galaxies or of the ratio of stellar to dark matter in them. It seems to me to support Kormendy's (this meeting) statement about a trend in M/L with luminosity, although it is clearly desirable to carry out a rather more detailed analysis of more than two systems.

One other interesting recent development in spherical galaxies has been DeJonghe's calculation of the distribution of projected velocities in a Plummer model. DeJonghe showed that the profile is bimodal if the dispersion tensor is very tangentially elongated. Below we illustrate this effect for power law light distributions in a logarithmic potential with the circular velocity everywhere unity. An isotropic distribution function would have a Gaussian $f(v_p)$ at all radii, with dispersion given by $\sigma^2 = v_c^2/K$, where $K = -d\ln\epsilon/d\ln r$. The observed profiles are decidedly not Gaussian. Further investigation is required to see if this effect can be observed after convolution with a stellar template. If so, it offers a possible approach to breaking the degeneracy between velocity dispersion and anisotropy in these models.

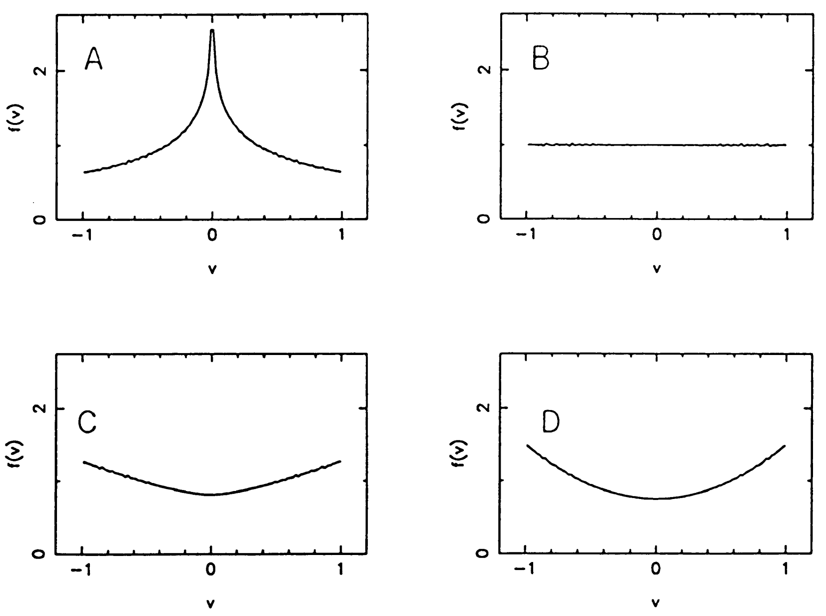

Figure 5. – Observed velocity distributions for power law models described in text. All models have $v_c = 1$. Values of K were as follows: A: 2, B: 3, C: 4, D: 5.

https://doi.org/10.1017/S0074180900185237 Published online by Cambridge University Press

REFERENCES

Binney, J. and Mamon, G. A. 1982, *Mon. Not. Roy. Astr. Soc.* **200**, 361.
DeJonghe, H. 1987, *Mon. Not. Roy. Astr. Soc.* **224**, 13.
Dressler, A. 1980, *Astroph. J. Letters*, **240**, l11.
Dressler, A. 1984, *Astroph. J.*, **286**, 97.
Duncan, M. J. and Wheeler, J. C. 1980, *Astroph. J.* **237**, L27.
Eddington, A. 1916, *Mon. Not. Roy. Astr. Soc.* **76**, 572.
Kent, S. M. and Gunn, J. E. 1982, *Astron. J.* **87**, 945.
Lawson, C.L. and Hanson, R.J. 1974, *Solving Least Squares Problems*, (Prentice Hall, Englewood Cliffs, New Jersey).
Merritt, D. 1985, *Astron. J.* **90**, 1027.
Michie, R. W. and Bodenheimer, P. H. 1963, *Mon. Not. Roy. Astr. Soc.* **126**, 269.
Newton, A. J. and Binney, J. 1984, *Mon. Not. Roy. Astr. Soc.* **210**, 711.
Pfenniger, D. 1984, *Astron. Astroph.* **141**, 171.
Richstone, D. O. and Tremaine, S. 1984, *Astroph. J.* **286**, 27.
Richstone, D. O. and Tremaine, S. 1985, *Astroph. J.* **296**, 370.
Richstone, D. O. and Tremaine, S. 1986, in preparation.
Sargent, W. L. W., Young, P. J., Boksenberg, A., Shortridge, K., Lynds, C. R., Hartwick, F. D. A., 1978, *Astroph. J.* **221**, 731.
Schwarzschild, M. 1979 *Astroph. J.* **232**, 236.
Tonry, J. 1983, *Astroph. J.* **266**, 58.
Tonry, J. 1984, *Astroph. J. Letters* **283**, L27.
Tremaine, S., Henon, M., and Lynden-Bell, D. 1986, *Mon. Not. Roy. Astr. Soc.* **219**, 285.
Vandervoort, P. O. 1984, *Astrophys. J.* **287**, 475.
Young, P. J., Sargent, W. L. W., Boksenberg, A., Lynds, C. R. and Hartwick, F. D. A. 1978, *Astroph. J.* **222**, 450.

DISCUSSION

Jaffe: Is looking for non–gaussian velocity profiles a useful way to resolve the ambiguities in solving for the distribution function?

Richstone: Dressler and I tried to do that for the center of M32 with a very extreme distribution function. After convolution with a standard star the observed line profile is nearly indistinguishable from a Gaussian convolved with a star. We were discouraged and did not proceed further. Dejonghe's work ***suggests*** that it is worthwhile to pursue this question at large radii.

Lupton: In view of the short relaxation time in the core of M32 ($\sim \frac{1}{40} H_0^{-1}$), does a one–component, constant M/L model make any sense?

Richstone: At $\sim 1''$ the relaxation time is about $\frac{1}{2} H_0^{-1}$. The traditional logic in searching for black holes or dark matter is to see first whether they are mandated by a failure to produce a constant M/L model. In this case, it seems to be impossible to fit the observed velocities with a constant M/L model.

https://doi.org/10.1017/S0074180900185237 Published online by Cambridge University Press

Kormendy: The core radius of M32 is $<< 1''$. At $r_c = 1.3$ pc, the best limit I can get from my observations, the relaxation time is $\sim \frac{1}{30}$ Hubble time. The relaxation time becomes equal to a Hubble time at about 2–3$''$ radius.

Burstein: Over what range of radii do your M/L–estimates for M32 and M87 apply?

Richstone: About 3$''$ to about 20$''$ in M32. 0$''$ to about 60$''$ in M87. In both cases M/L was assumed to be constant over those ranges.

King: The M/L that is derived is not global. It will apply to the region where the velocities are observed.

Binney: A short sermon on entropy. E. T. Jaynes (*e.g. Papers on probability, statistics and statistical physics*, ed. R. D. Rosenkrantz, Reidel 1983; also Dejonghe, H., 1987, *Astroph. J.*, in press) argues eloquently that the Gibbs–Boltzmann entropy $S = -\sum_i p_i \ln p_i$ enjoys a special place amongst the many convex functionals that can be used for rating probability distributions p_i. Jaynes contends that S is a purely subjective quantity that enables us to decide which of two probability distributions, that are both compatible with the available data, is more plausible. The structure of S is enforced by the laws of probability theory, and *has nothing whatever to do with physics.* In particular, the additivity of information on which Shannon's uniqueness proof rests, is not connected with the kind of physical additivity that arises in extensive thermodynamic systems. Shannon's additivity is to do with the variation with n of amount of information communicated by telling someone how many stars are in each of a series of phase-space cells of size $n\tau$. We obtain the most plausible probability distribution $p_i \equiv f(w_i)d^6w$ (and the only consistent interpretation of the distribution function is as a *probability* density—attempts to interpret f as some kind of stellar *density* generate only confusion) by maximizing S subject to *all* available constraints. As Jaynes emphasizes, attacks on the maximum-entropy procedure generally consist in obtaining manifestly absurd results by failing to include an important piece of prior information in the constraints, and we should beware of falling into this trap.
We still do not know what are the essential constraints for stellar systems. Usually people impose values of E and M, and sometimes a constraint derived from a theory of galaxy formation, such as the maximum phase-space density. One can think of many other constraints, e.g., the value of a tidal radius or some kinematic data (see also Dejonghe, *ibid.*). The consistent application of the maximum entropy procedure would involve repeatedly maximizing S subject to an ever-lengthening list of constraints, until further constraints do not significantly improve the agreement with observation.
Finally, let me remark that the following mathematical fact is an important source of confusion. If f_n maximizes S subject to constraints $C_1, \ldots, C_n$, then the distribution f_{n+1} that maximizes S subject to $C_1, \ldots, C_{n+1}$ is the same as the distribution f that maximizes $S' \equiv -\int d^6w f(w) \ln[f(w)/f_n(w)]$ subject to the single constraint C_{n+1}. In other words, one can consolidate a long list of constraints into a "prior" f_n. Later, we tend to forget about the constraints $C_1, \ldots, C_n$ and imagine that we are simply maximizing a new entropy S' subject to a single constraint. There is no objection to this way of thinking when S' *has* been properly founded at an earlier stage, but we should not listen to people who pull "entropies", or H–functions (Tremaine *et al.* *M.N.R.A.S.* **219**, 285), out of thin air, and claim that they enjoy the same status as S.

Part of the Dutch connection: Reynier Peletier (with his back to the camera), Andrew Pickles, Peter Teuben, Marijn Franx & Myriam Hunink.