SOME CONSIDERATIONS IN MAKING FULL USE OF
THE HIPPARCOS CATALOGUE

A.G.A. BROWN[1], F. ARENOU[2], F. VAN LEEUWEN[3], L. LINDEGREN[4], X. LURI[2,5]
[1] *Sterrewacht Leiden, Postbus 9513, 2300 RA Leiden, The Netherlands*
[2] *Observatoire de Paris, Section d'Astrophysique, 92195 Meudon Cedex, France*
[3] *Royal Greenwich Observatory, Madingley Road, Cambridge, CB3 0EZ, United Kingdom*
[4] *Lund Observatory, Box 43, S-22100, Lund, Sweden*
[5] *Departament d'Astronomia i Meteorologia UB, Avda. Diagonal 647,*
E08028 Barcelona, Spain

Introduction.   This contribution is intended as a 'rough guide' to the Hipparcos Catalogue for the non-expert user. We shall focus here on the interpretation of the measured parallaxes in terms of distances and luminosities of stars, and on the use of the covariance matrix of the errors on the astrometric parameters. Issues not addressed here are the level at which one may expect systematic errors to occur in the Hipparcos Catalogue, and the correlation of astrometric parameters over a given region of the sky. For more details on the latter two subjects we refer to the documentation of the Hipparcos Catalogue (ESA, 1997) and to Brown *et al.* (1997). We note that at present the Hipparcos Catalogue contains no identified systematic errors.

Correct Use of Trigonometric Parallaxes.   Notwithstanding the unprecedented quality of the Hipparcos data, the correctness of the *astrophysical* results is not assured, as the estimation of stellar distances, absolute magnitudes and other physical quantities from trigonometric parallaxes is not a trivial process. The statistical properties of the relationships involved and the effects of sample selection hide several pitfalls that, if not avoided, lead to biased estimates. In the following π and π_H stand for the true and measured parallax of a star.

Several quantities, such as the stellar distance or the absolute magnitude, have a non-linear dependence, $h(\pi)$, on the parallax. In this case, the expectation value of the function, $E[h(\pi_H)]$, is in general different from $h(\pi)$, even if the individual Hipparcos parallaxes are unbiased, i.e. if $E[\pi_H] \approx \pi$. In other words, $1/\pi_H$ is a biased estimate of the star's true distance, and $m+5\log\pi_H+5$ is a biased estimate of its absolute magnitude: $E[1/\pi_H] \neq 1/\pi$ and $E[m+5\log\pi_H+5] \neq m+5\log\pi+5$. The expectation values of distance and absolute magnitude can be calculated. For example, if parallaxes below 0.01 mas, which is the smallest non-zero value of parallax that can be found in the catalogue, are excluded the computed distance may be overestimated by more than 100 per cent when the relative error in π is 100 per cent. Note that if negative parallaxes are not rejected, the bias, although reduced, is still present. Absolute magnitudes computed from observed parallaxes are on the average 0.2 mag too bright when the relative parallax error is about 50 per cent, and 0.6 mag too faint for a 200 per cent relative error.

Both bias and variance are negligible for relative errors better than about 10 per cent. For parallaxes with a higher relative error, one could naively hope to correct the computed distance from the calculated bias, but this is not possible because the bias is a function of σ_{π_H}/π, where the *real* parallax π is unknown. What is available is *not* the real relative parallax error but the observed one σ_{π_H}/π_H. Given the uncertainty on the true relative error, a bias correction is simply not feasible. On the other hand, such a correction would only have a statistical meaning when applied to a sample, but would be questionable on an individual basis.

Another source of biases are the selection effects in a sample. A well known selection bias is the Malmquist bias. The use of the parallaxes of a truncated sample without caution may lead to similar biases in the derived results. Another type of bias is caused by an indirect truncation of the parallaxes. For instance, suppose that the spatial velocities of a given sample are computed

547

and stars with high spatial velocity are selected. This will select stars with a truly high velocity but also stars with an overestimated distance ($\pi_H \ll \pi$): the estimated distances of objects in this subsample will be biased in the mean. Consequently its estimated mean absolute magnitude will be too bright.

Summarizing, one can easily calculate the expected biases for a *given* true parallax. However, one only has the observed values, so the correction will depend on what kind of assumption one makes concerning the true values. In other words, the *distribution* of the true parallaxes has to be known, and this is an astrophysical question, not a statistical one! Hence, one cannot solve this problem just by statistics, but needs also some kind of modeling of the objects or sample under study. The reader is strongly encouraged to perform a detailed analysis of this sort *for each specific case* in order to obtain a correct estimation of any parameter of a star or a sample of stars using trigonometric parallaxes. This means in particular that one should neither ignore the possible biases nor apply blindly 'Malmquist' or 'Lutz-Kelker' corrections.

Use of the Covariance Matrix. One of the unique features of the Hipparcos Catalogue is that not only the standard errors of the five astrometric parameters are provided but also their correlation coefficients. This allows the user to make full use of the information contained in the astrometric parameters. Here we concentrate on using the covariance matrix when interpreting the statistics of a particular data set.

If $\mathbf{x}$ is an observed vector with covariance matrix $\mathbf{C_x}$ then the confidence region around $\mathbf{x}$ is given by: $c = \mathbf{x}'\mathbf{C_x^{-1}}\mathbf{x}$, where the prime denotes matrix transposition. The distribution of c is described by a χ_ν^2 probability distribution, where ν, the number of degrees of freedom, is equal to the dimension of $\mathbf{x}$. In the one-dimensional case this reduces to the well-known Gaussian distribution, where $c = 9$ corresponds to '3σ', the 99.73 per cent confidence level. For other values of ν the value of c will be higher for the same confidence level. It is 11.8 for $\nu = 2$ and 14.2 for $\nu = 3$. Note that the distribution of the errors around $\mathbf{x}$ is described by a multi-dimensional Gaussian and the equation above describes a confidence 'ellipsoid' around $\mathbf{x}$. If the vector $\mathbf{y}$ is derived from $\mathbf{x}$ via some transformation $\mathbf{f}(\mathbf{x})$, the covariance matrix of $\mathbf{y}$ is: $\mathbf{C_y} = \mathbf{J}\mathbf{C_x}\mathbf{J}'$. Here $\mathbf{J}$ is the Jacobian matrix of the transformation from $\mathbf{x}$ to $\mathbf{y}$: $[\mathbf{J}]_{ij} = \partial f_i / \partial x_j$. Thus one can calculate the covariance matrix of any set of variables derived from the observed astrometric parameters.

We now turn to the example of space velocities for cluster stars, specifically the Hyades. For full details we refer the reader to Perryman *et al.* (1997). When deriving space velocities for cluster stars we make use of the observed vector $(\pi, \mu_{\alpha*}, \mu_\delta, V_R)$, where $\mu_{\alpha*} = \mu_\alpha \cos\delta$, and V_R is the radial velocity. This vector is transformed to a space velocity, implicitly invoking a transformation to $(V_{\alpha*}, V_\delta, V_R)$ $(V_{\alpha*} = \mu_{\alpha*}A_v/\pi, V_\delta = \mu_\delta A_v/\pi, A_v = 4.74047... \text{ km yr s}^{-1})$. One can show that even in the absence of correlations between astrometric errors, the parallaxes and velocity components $V_{\alpha*}$ and V_δ will in general be correlated. Moreover, because of the position of the convergent point of the Hyades the correlated errors will lead to systematic behaviour of the uncertainties in the sample as a whole. These systematics will be transferred to the space velocities.

In the case of the Hyades both the overall distribution of the velocity residuals (with respect to the mean cluster motion), as well as the correlation of the direction of the residuals with spatial position, can be fully attributed to observational errors. We stress here that ignoring the covariance matrix can easily lead to false interpretation of, for example, kinematic data. For cases other than the Hyades the way in which features due to correlated errors enter may differ. It is important to carry out this kind of analysis and consider the implications for each case individually.

Finally, we refer again to Brown *et al.* (1997) and Perryman *et al.* (1997) for many more details and further references than can be provided here.

References

ESA, 1997, *The Hipparcos and Tycho Catalogues*, ESA SP-1200
Brown, A.G.A., Arenou, F., Lindegren, L., van Leeuwen, F., Luri, X., 1997, in *Hipparcos Venice '97*, ESA SP-402
Perryman, M.A.C., Brown, A.G.A., Lebreton, Y., Gómez, A., Turon, C., Cayrel, G., Mermilliod, J.-C., Robichon, N., Kovalevsky, J., Crifo, F., 1997, *Astron. & Astroph.*, in press