A COMPARATIVE ANALYSIS OF SOME STATISTICAL PROPERTIES OF THE HIPPARCOS CATALOGUE

B. Bucciarelli

Osservatorio Astronomico di Torino, 10025 Pino Torinese TO, Italy

ABSTRACT

Due to the unprecedented astrometric quality of the Hipparcos Catalogue, presently available astrometric catalogues cannot be used to independently assess its statistical properties. However, a relatively inexpensive way to address systematic errors at the sub-mas level is to intercompare the FAST and NDAC results before their merging. Moreover, by means of an empirical covariance analysis it is possible to study the presence of residual signals in the Hipparcos sphere solution. This papers presents some results of these investigations.

Key words: astrometry; reference catalogues; statistics.

1. SYSTEMATIC ERRORS ANALYSIS

1.1. Catalogue-to-Catalogue Reduction

Prior to this analysis, the FAST and NDAC catalogues have been put onto the same reference system (H30, i.e. approximately FK5). Then, the method of infinitely overlapping circles (IOC) has been used to analyse the high angular frequency systematics, following the prescriptions of a similar study performed on the FAST/NDAC 30-month solutions (Kovalevsky et al. 1995). The IOC procedure (Bucciarelli et al. 1994 and references therein), briefly consists of a generalized moving mean algorithm used to find an optimal *weight* for each star in order to evaluate the local systematic differences between two different catalogues. For each star, the formula to compute such weight reads:

$$\omega(\rho) = (\frac{2}{\pi})[\cos^{-1}(\frac{\rho}{r}) - (\frac{\rho}{r})[1-(\frac{\rho}{r})^2]^{\frac{1}{2}}] \quad 0 \le \rho \le r$$

where r is the radius at which the correlation is practically dropping to zero, and ρ is the distance between the central star and the star being weighted. By using such a definition of statistical weight, which does include the central star, one naturally generates continuous systematic differences, while still treating the random part of the individual residuals in a statistically correct way, i.e. the formal expectation of the random part is still zero.

The radius of the small circle has been set to 2 degrees, giving an average of about 30 stars per IOC circle. This choice was driven by the request that the influence of random errors be minimized, while still probing small scale systematics. This instance is crucial, since the random error of the astrometric parameters is of the same order of magnitude, and even larger, than the systematic effects under investigation. This technique has been applied to all five astrometric parameters as a function of position on the celestial sphere. The data have been binned in 90 longitude wedges and 2-degree wide latitude strips. As expected, the values are small, typically of the order of, or smaller then, 0.1 mas. Figures 1(a), (b) and (c) show the graphs obtained as a function of ecliptic latitude. The reason for the large discrepancy in one bin of Figure 1(a) has not been further investigated, and is most likely due to an object mismatch.

1.2. Testing the Empirical Distribution Functions

Once the residual systematics have been removed from either one of the two catalogues, we have computed the empirical distribution function of the normalized differences between FAST and NDAC, and compared it to the theoretical one, as explained below.

In right ascension the test statistic is (and analogously for any of the other astrometric parameters):

$$\epsilon_\lambda = \frac{|\lambda_f - [\lambda_n - \Delta\lambda_{f,n}]|}{[m^2(\lambda_n) + M^2(\lambda_f) - 2\times\rho_{n,f}\times m(\lambda_n)M(\lambda_f)]^{\frac{1}{2}}}$$

where $m^2(\lambda_n)$ is the variance of the NDAC right ascension, $M^2(\lambda_f)$ the variance of the FAST right ascension and $\rho_{n,f}$ their correlation coefficient. The quantity $\Delta\lambda_{f,n}$ is the catalogue-to-catalogue systematic difference derived with the IOC averaging technique. The predicted distribution for ϵ_λ (or any of the other astrometric parameters) is a folded Gaussian with a mean of 0.798 and a rms of 0.603. The actual values of the first two moments of the distributions are in good agreement with the theoretical expectations. Note that the distribution of ϵ is degenerate for the case of complete overlap between the two catalogues. However, we used a correlation coefficent of 0.79, a reduction of about 15 per cent with respect to its estimated nominal value (~ 0.9), to

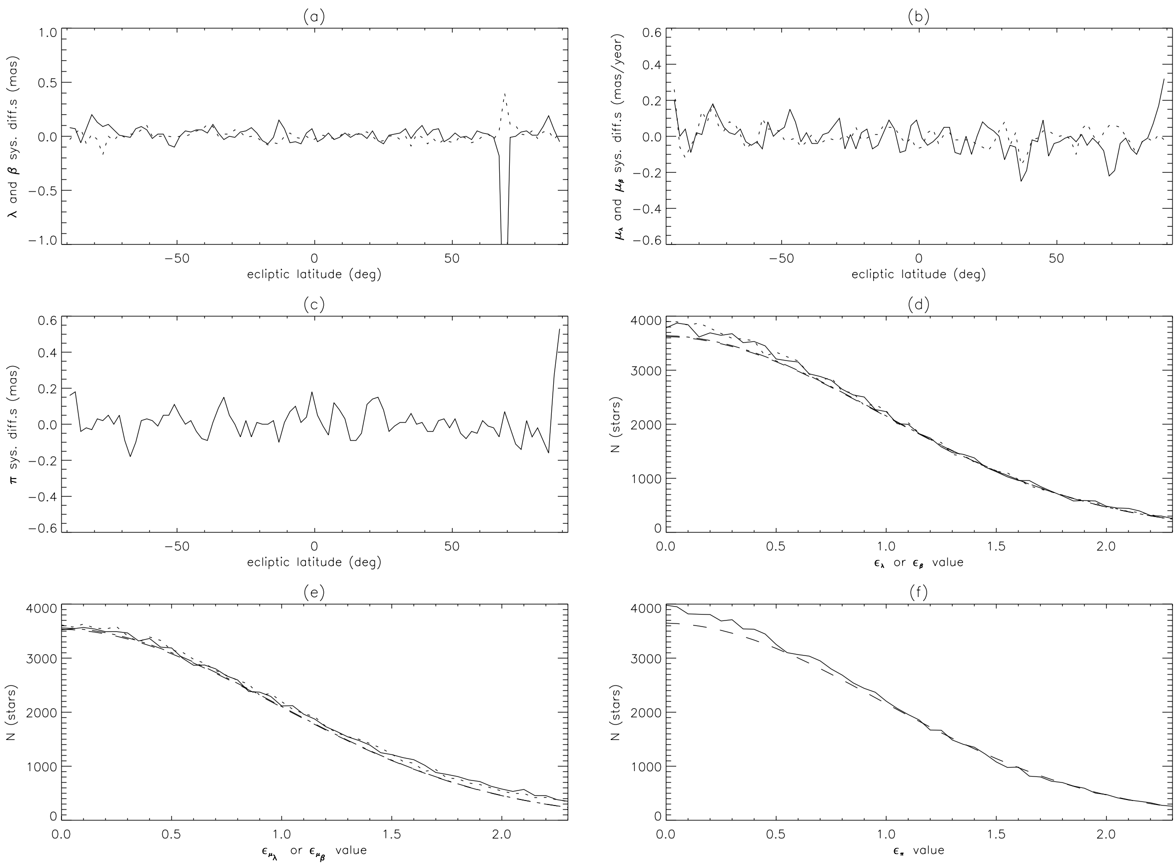

Figure 1. Diagrams (a), (b), and (c) are plots of the systematic differences between FAST and NDAC as a function of ecliptic latitude; graphs (d), (e), and (f) show the empirical distribution functions of the ϵ statistic against the theoretical (dashed lines) ones, for a catalogue-to-catalogue correlation coefficient of 0.79 (solid and dotted lines refer to longitude and latitude respectively when either position or proper motion components are combined).

obtain the results discussed below. Part of the foundation for such a diminishment of the catalogue-wise correlation coefficient lies in the different processing paths adopted by the two consortia, which differentiate the catalogues more than what the number of common observations would suggest, thereby lifting (in practice) the apparent 'degeneracy' of the problem.

Figures 1(d), (e) and (f) show the empirical distribution functions ϵ_λ, ϵ_β, $\epsilon_{\mu_\lambda cos\beta}$, ϵ_{μ_β}, and ϵ_π and their theoretical counterparts. Dashed lines are used for the theoretical curves, while dotted and solid lines represent the observed ones. Mean and standard deviation values of the empirical distributions are reported in Table 1.

In all cases a relatively small number of 'outliers' ($\sim$ 3 per cent or less) were found, which have not been taken into account in the calculation of the expectation values. The presence of these outliers is usually explained as a discrepancy between the actual differences and the formal errors listed in the catalogues.

In conclusion, this analysis shows that the level of (internal) residual systematics is at the level of, or smaller than, that expected from pre-launch estimates. Also, the formal errors, as tested by the ϵ distributions, appear to have a high degree of consistency with statistical theory.

Table 1. Sample mean values and standard devations for the empirical distribution functions. The theoretical values are 0.798 and 0.603 respectively.

Statistic	Mean value	Standard deviation
λ	0.769	0.565
β	762	0.561
$\mu_\lambda cos\beta$	0.808	0.588
μ_β	0.798	0.581
π	0.759	0.560

2. EMPIRICAL COVARIANCE ANALYSIS

Following the prescriptions of Sansó (1990), the following analysis has been carried out on the 37-months, non-iterated, FAST sphere solution. Let ξ represent the adjustment to the astrometric parameter of a generic star S output of the Hipparcos sphere solution. To study the statistical properties of this so-

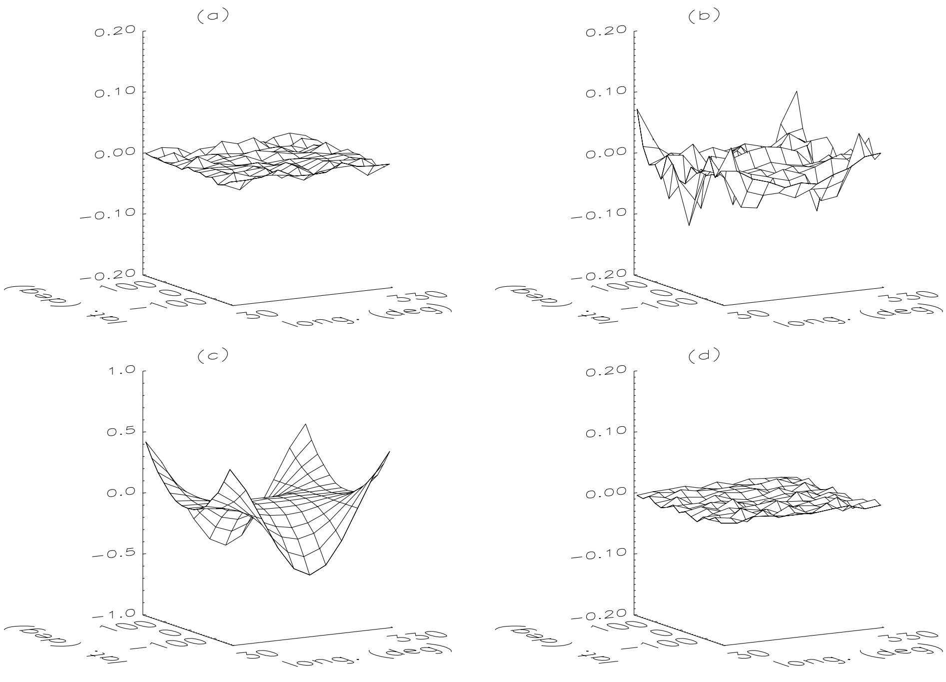

Figure 2. Surface plots of the empirical covariance as computed for a pseudo-random variate (a), and for $\delta\lambda cos\beta$ (b), $\delta\mu_{\lambda*}$ (c), and $\delta\pi$ (d)

lution one can apply a continuous (empirical) covariance analysis to model the correlations of a chosen stellar parameter as function of, e.g. ecliptic latitude (λ). If such function exists, then $\xi(\lambda)$ must be periodic of period 2π, i.e:

$$\xi(\lambda) = C_{\circ\xi} + \sum_{n=1}^{\infty}(C_{n\xi}\cos(n\lambda) + S_{n\xi}\sin(n\lambda)) \quad (1)$$

From the definition $C_{\xi\xi} \equiv E[\xi(\lambda+\Delta\lambda)(\xi(\lambda)] - C_{0\xi}^2$, after some calculations one obtains:

$$C_{\xi\xi}(\Delta\lambda) = \sum_{n=1}^{\infty}(\frac{C_{n\xi}^2}{2} + \frac{S_{n\xi}^2}{2})\cos(n\Delta\lambda) \quad (2)$$

Equation 2 is the covariance function to be estimated; in practice, the series must be truncated to a suitable N (e.g. $N = 30$). From Fourier theory, the following expressions hold for coefficients $C_{n\xi}$ and $S_{n\xi}$:

$$C_{n\xi} = \frac{1}{\pi}\int_0^{2\pi}\xi(\lambda)\cos(n\lambda)d\lambda \quad (3)$$

$$S_{n\xi} = \frac{1}{\pi}\int_0^{2\pi}\xi(\lambda)\sin(n\lambda)d\lambda \quad (4)$$

To compute such coefficients, the integrals in Equationa 3 and 4 must be discretized as:

$$\hat{C}_{n\xi} = \frac{1}{\pi}\sum_{i=1}^{M}\xi_{\rm oss}(\lambda_i)\Delta_i\cos(n\lambda_i) \quad (5)$$

where M is the total number of stars, $\xi_{\rm oss}$ is the vector of least-squares estimates of the adjustments to the stars' positions, and Δ_i represents the finite integration step, which depends on the actual stellar distribution. Analogous formula holds for $\hat{S}_{n\xi}$. After computing the Fourier coefficients, one has to accept them as significanlty different from 0. To test their significance, let us proceed as follows. Let us write:

$$\xi_{\rm oss}(\lambda) = \xi(\lambda) + \epsilon \quad (6)$$

where ϵ is a purely random noise error component, from Equation 5 one obtains:

$$\sigma^2(\hat{C}_{n\xi}) = \frac{\sigma_\epsilon^2}{\pi^2}\sum_{i=1}^{M}\Delta_i^2\cos^2(n\lambda_i) \quad (7)$$

An estimate of σ_ϵ^2 is calculated from (6) and (2), noticing that $C_{\xi\xi}(\Delta\lambda = 0) = \sigma^2(\xi)$, as:

$$\hat{\sigma_\epsilon}^2 = \hat{\sigma}^2(\xi_{\rm oss}) - \sum_{n=1}^{N}(\hat{C}_{n\xi}^2/2 + \hat{S}_{n\xi}^2/2) \quad (8)$$

where $\hat{\sigma}^2(\xi_{\rm oss})$ comes from the least-squares solution.

If the expectation value of $\hat{C}_{n\xi}$ is equal to zero, then the Central Limit Theorem states that variable $\hat{C}_{n\xi}/\sigma(\hat{C}_{n\xi})$ has a standard probability distribution. Then a 95 per cent level confidence interval has been used to test the hypothesis that $\hat{C}_{n\xi}$ is not different from zero. Proceeding in such a way for each coefficient included in the Fourier series we have computed $C_{\xi\xi}(\Delta\lambda)$ from Equation 2, retaining only those coefficients that have resulted significantly different from zero. Finally, the estimated covariance was normalized by dividing it by the mean square error of the pertinent ξ.

This method has been applied to the 37-months, non-iterated, FAST sphere solution obtained using only primary, i.e. single, well-behaved, stars, for a total of 42 035 stars, and for different declination strips. Since the number of stars per declination strip is obviously critical to the computation of the Fourier coefficients, we have divided the sphere in 15 equal-area latitude strips to obtain statistical samples of comparable sizes. The sampling of the empirical covariance was chosen to be $\Delta\lambda \sim 30°$, which corresponds approximately to half the basic angle defined by the Hipparcos complex mirror. In order to check the software, a set of adjustments to the real stellar distribution was simulated using pseudo-random numbers. When the same covariance analysis is applied to this sythetic solution, the covariance terms turn out to be practically zero, as shown in Figure 2(a).

Figures 2(b), (c), and (d) are surface plots of the functions $C_{\delta\lambda*\delta\lambda*}(\Delta\lambda)$, $C_{\mu_{\lambda*}\mu_{\lambda*}}(\Delta\lambda)$, and $C_{\delta\pi\delta\pi}(\Delta\lambda)$ respectively (the latitude components exhibit similar behaviours). It can be seen that there is no significant correlation term in position, and that the correlation of parallax is negligible; on the contrary, the correlation surface of the proper motion parameter shows a more definite pattern, which deserves further investigation.

REFERENCES

Bucciarelli B., Lattanzi M.G. , Taff L.G., 1994, ApJ, 433, 831

Kovalevsky, J., Lindegren, L., Froeschlé, M., van Leeuwen, F., Perryman, M.A.C., Falin, J.L., Mignard, F., Penston, M.J., Petersen, C., Bernacca, P.L., Bucciarelli, B., Donati, F., Hering, R., Høg, E., Lattanzi, M., van der Marel, H., Schrijver, H., Walter, H.G., 1995, A&A, 304 , 34–43

Sansó, F. 1990, private communication.