Relativity in Fundamental Astronomy
Proceedings IAU Symposium No. 261, 2009
S. A. Klioner, P. K. Seidelman & M. H. Soffel, eds.

doi:10.1017/S1743921309990573

Spatial correlations in the Gaia astrometric solution

Berry Holl[1], **David Hobbs**[1] **and Lennart Lindegren**[1]

[1]Lund Observatory, Lund University, Box 43, SE-22100 Lund, Sweden
email: berry@astro.lu.se

Abstract. Accurate characterization of the astrometric errors in the forthcoming Gaia catalogue is essential for making optimal use of the data. Using small-scale numerical simulations of the astrometric solution, we investigate the expected spatial correlation between the astrometric errors of stars as function of their angular separation. Extrapolating to the full-scale solution for the final Gaia catalogue, we find that the expected correlations are generally very small, but could reach some fraction of a percent for angular separations smaller than about one degree. The spatial correlation length is related to the size of the field of view of Gaia, while the maximum correlation coefficient is related to the mean number of stars present in the field at any time. Our scalable simulation tool (AGISLab) makes it possible to characterize the astrometric errors and correlations, e.g., as functions of position and magnitude.

Keywords. Astrometry, reference systems, catalogs, methods: data analysis, methods: statistical, space vehicles

1. Introduction

The space astrometry mission Gaia, planned to be launched in 2012 by the European Space Agency (ESA), will provide the most comprehensive and accurate catalogue of astrometric data for galactic and astrophysical research in the coming decades. Accuracies of 8–25 μas are expected for the trigonometric parallaxes, positions at mean epoch and annual proper motions of simple (apparently single) stars down to 15th magnitude with lower accuracy down to 20th magnitude. The astrometric data are complemented by photometric and spectroscopic information collected by the satellite. The resulting catalogue will become available to the scientific community around 2020.

Accurate characterization of the errors in the catalogue is essential for making optimal use of the data. While estimates of the standard errors for individual stars have been reported elsewhere (e.g., Lindegren 2009), little is yet known about what happens when Gaia data are combined for large numbers of objects. This will often be the case in important applications dealing with stellar clusters, nearby dwarf galaxies, galactic stellar populations, and when looking for large-scale patterns e.g. in the apparent proper motions of quasars. In such cases it may be important to know the statistical correlation of the astrometric errors as a function of the angular separation of the objects, which we refer to as *spatial correlations*. In this paper we present the first results of an estimation of spatial correlations in the future Gaia catalogue.

2. The importance of correlations

We examine the estimated value of the generic astrometric parameter x (representing α, δ, π, μ_α, or μ_δ) for star i, denoted x_i. It is assumed that the estimate is unbiased, so $\mathrm{E}[e_i] = 0$, where $e_i = x_i - x_i^{\mathrm{true}}$ is the error. For the two stars $i \neq j$ the estimates are correlated with correlation coefficient ρ_{ij} if

$$\mathrm{Cov}[x_i, x_j] \equiv \mathrm{E}[e_i e_j] = \rho_{ij}\sigma_i\sigma_j \neq 0 \tag{2.1}$$

where $\sigma_i = \sqrt{\mathrm{E}[e_i^2]}$, $\sigma_j = \sqrt{\mathrm{E}[e_j^2]}$ are the standard errors.

Consider now any quantity y calculated from the estimated parameters $x_1 \dots x_N$ of N different stars. We can generally formulate this as $y = f(\boldsymbol{x})$ where $\boldsymbol{x}$ is the vector of estimates. Assuming that f is linear in the small errors, the variance of y is given by

$$\begin{aligned}\sigma_y^2 &= \left(\frac{\partial y}{\partial \boldsymbol{x}}\right)' \mathrm{Cov}(\boldsymbol{x}) \left(\frac{\partial y}{\partial \boldsymbol{x}}\right) \\ &= \sum_i \left(\frac{\partial y}{\partial x_i}\right)^2 \sigma_i^2 + \sum_i \sum_{j\neq i} \frac{\partial y}{\partial x_i}\,\frac{\partial y}{\partial x_j}\,\rho_{ij}\sigma_i\sigma_j \end{aligned} \tag{2.2}$$

The first sum is the computed variance if correlations are neglected; depending on the sign of the correlations this can be an under- or overestimate of the true variance.

As a simple example, consider the calculation of the mean parallax or proper motion of N stars in a cluster, so $y = N^{-1}\sum_i x_i$. If the stars are of approximately the same magnitude, they will have roughly the same standard error, $\sigma_i \simeq \sigma$. If the area on the sky occupied by the cluster is small, it will be found (cf. Sect. 5) that the correlation coefficient is positive and roughly the same for all pairs of stars, $\rho_{ij} \simeq \rho > 0$. Then

$$\sigma_y^2 \simeq \sigma^2 \left(\frac{1}{N} + \frac{N-1}{N}\rho\right) \tag{2.3}$$

In the absence of correlations the improvement in σ_y is by a factor $N^{-1/2}$, as could be expected. However, in the presence of (positive) correlations, $\sigma_y \to \sigma\sqrt{\rho}$ as $N \to \infty$. The limiting accuracy is effectively reached by averaging over some ρ^{-1} stars.

3. Origin of spatial correlations in the Gaia catalog

Although the individual positional measurements in Gaia's focal plane are essentially uncorrelated (the errors are dominated by photon noise), the geometry of the observations and the way they are combined in the astrometric solution will create spatial correlations on different angular scales. Gaia's scanning law and the two fields of view, widely separated by the basic angle of 106.5°, are designed to minimize large-scale correlations, but cannot entirely eliminate them.

The origin of spatial correlations in the Gaia catalog can be understood in terms of errors in the attitude determination (Fig. 1). An error in the attitude at a particular time will 'bias' all observations made at that time, in both fields of view, thus partially correlating stars within each field of view as well as stars separated by the basic angle. This suggests that we can expect spatial correlations to fall off over an angular scale on the order of the field of view size, or $\simeq$0.7° for the astrometric instrument of Gaia, possibly reappearing to some degree for separations of about 106.5°.

In a single realization of the Gaia catalogue (which is all that we will have!), the spatially correlated errors may look like localized 'biases' on the sky. However, it is important to realize that they originate from attitude errors that are themselves also a result of the random observation noise and that a different realization of the observation noise would have resulted in a completely different set of 'biases'.

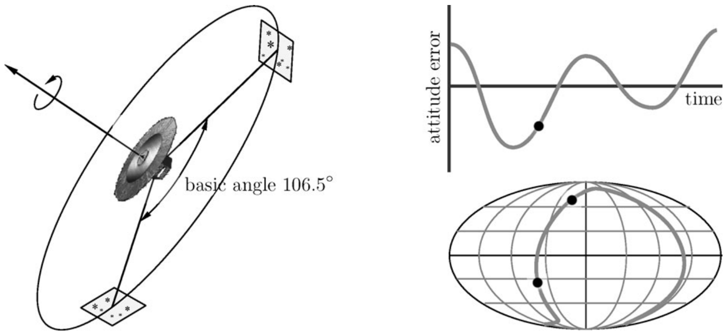

Figure 1. Gaia scans the sky roughly along great circles, observing stars in two fields of view separated by the basic angle. An error in the attitude at a particular time will affect all observations in both fields of view, producing correlations among stars both for small angular separations and for separations of about 106.5°, as illustrated by the all-sky map on the lower right.

4. Monte Carlo experiments with AGISLab

The baseline method for determining the astrometric parameters of Gaia stars is the Astrometric Global Iterative Solution (AGIS). This is an iterative least-squares estimation of the five astrometric parameters for a subset of $\sim 10^8$ well behaved (non-variable, apparently single) 'primary' stars, with additional unknowns for the spacecraft attitude, instrument calibration, and global parameters such as PPN γ (Hobbs *et al.* 2009). The total number of unknowns is $\sim 5 \times 10^8$. This large number prevents a rigorous calculation of the covariance matrix of the solution. To overcome this we estimate the correlations statistically from Monte Carlo experiments with different noise realizations.

While AGIS is currently being tested with 10^6–10^7 primary stars, these simulations take too much time and resources for making a significant number of Monte Carlo experiments. They also depend on externally generated simulated observations which are not always suitable for the tests we want to run. We have therefore developed a scaled version of AGIS called AGISLab, which allows us to run simulations with less than 10^6 stars in a (much) shorter time and using input observations that fit our experiments (e.g., with many different noise realizations but otherwise identical conditions). The scaling uses a single parameter S such that $S = 1$ corresponds to the astrometric solution using approximately the current Gaia design and a minimum of 10^6 primary stars, while $S = 0.1$ would only use 10% as many primary stars. When $S < 1$ the Gaia design used in the simulations is modified to preserve certain key quantities such as the mean number of stars in the focal plane at any time, the mean number of field transits of a given star over the mission, and the mean number of observations per degree of freedom of the attitude model. In practice this is done by formally reducing the focal length of the astrometric telescope and the spin rate of the satellite by the factor $S^{1/2}$, and increasing the interval between attitude spline knots by the factor S^{-1}.

For the present study we made astrometric solutions with AGISLab, using 3 000, 10 000, and 30 000 uniformly distributed primary stars (i.e., for $S = 0.003$, 0.01 and 0.03). In each experiment (A, B, C) many different noise realizations were made and the corresponding solutions computed in order to improve the statistics. As the scaling preserves the mean number of stars per field of view, additional experiments (D, E) were made in which this number could be increased. An overview of the experiments is given in Table 1. All experiments used a noise level of 100 μas per along-scan observation, which is roughly the expected noise for bright stars down to magnitude $V = 13$ (for unreddened G2V stars).

https://doi.org/10.1017/S1743921309990573 Published online by Cambridge University Press

Table 1. Overview of experiments in this study. S = scale parameter; $N_{\rm star}$ = number of stars in the solution; Φ = field of view size (side length); n = mean number of stars per field of view; $N_{\rm run}$ = number of runs in the experiment (with different noise realizations); $\rho_{\rm max}$ = maximum correlation of parallaxes (for separations $\theta \ll \Phi$); $\theta_{1/2}$ = correlation half-length. The reference case for the scaling law (corresponding to $S = 1$) has $N_{\rm star} = 10^6$, $\Phi = 0.7°$ and $n = 12$.

Experiment	S	$N_{\rm star}$	Φ [deg]	n	$N_{\rm run}$	$\rho_{\rm max}$	$\theta_{1/2}$ [deg]	Figure
A	0.030	30 000	4	12	49	0.085	1.7	2, 3
B	0.010	10 000	7	12	112	0.085	3.0	2, 3
C	0.003	3 000	13	12	759	0.085	5.2	2, 3, 4
D	0.003[1]	9 000	13	36	25	0.032	5.2	4
E	0.003[1]	30 000	13	120	17	0.010	5.0	4

[1] In these experiments n is increased with respect to the usual scaling law.

Although the actual noise level is irrelevant for studying correlations, the assumption of a single noise level, as well as the uniform sky distribution of the stars, are of course gross simplifications of the real case. These complications will be addressed in a future paper.

To estimate the correlation as a function of pair separation θ, the range $0 \leqslant \theta \leqslant 180°$ was divided into bins of 0.5° or 1.0° and the sample correlation coefficient calculated in each bin by summing over all relevant pairs in all the runs:

$$\rho(\theta) = \left(\sum_{ij\in\text{bin}} e_i e_j\right)\left[\left(\sum_{ij\in\text{bin}} e_i^2\right)\left(\sum_{ij\in\text{bin}} e_j^2\right)\right]^{-1/2} \tag{4.1}$$

5. Results and discussion

Results are shown graphically in Figs. 2–4. In all cases the strongest correlation is obtained for the smallest separations. For $\theta \simeq 106.5°$ there is also a much weaker positive correlation for α and μ_α, and a similar negative correlation for π, δ and μ_δ (Fig. 2). This is expected if the attitude errors are mainly a rotation offset around the spin axis.

Table 1 gives two key characteristics of the correlation curves for small separations: the maximum correlation coefficient $\rho_{\rm max}$ (obtained in the first 0.5° bin), and the correlation half-length, i.e., the angle $\theta_{1/2}$ such that $\rho(\theta_{1/2}) \simeq \rho_{\rm max}/2$. It is noted that: (i) the correlation length scales with the size of the field of view, $\theta_{1/2} \simeq 0.4\Phi$; and (ii) the maximum correlation depends mainly on the number of stars in the field.

Extrapolating to the real Gaia mission with $\Phi = 0.7°$, the expected correlation half-length is $\simeq 0.3°$. The maximum correlation depends on the assumed number of primary stars and their magnitude distribution. Although the final solution will use about 100 million primary stars, their combined astrometric weight corresponds to a smaller number of perhaps 20 million bright primary stars, suggesting $\rho_{\rm max} \simeq 0.005$ for bright stars ($V < 13$) and smaller for fainter stars. More detailed studies are needed to determine how the magnitude distribution and real-sky non-uniformity affect the correlations. The final goal is to model the covariance of all astrometric parameters for any pair of stars in terms of their magnitudes and positions on the sky, as required for astrophysical applications combining Gaia data for many stars.

Acknowledgements. BH is an ELSA Fellow supported by the Marie Curie FP6 contract MRTN-CT-2006-033481. DH and LL acknowledge support by the Swedish National Space Board.

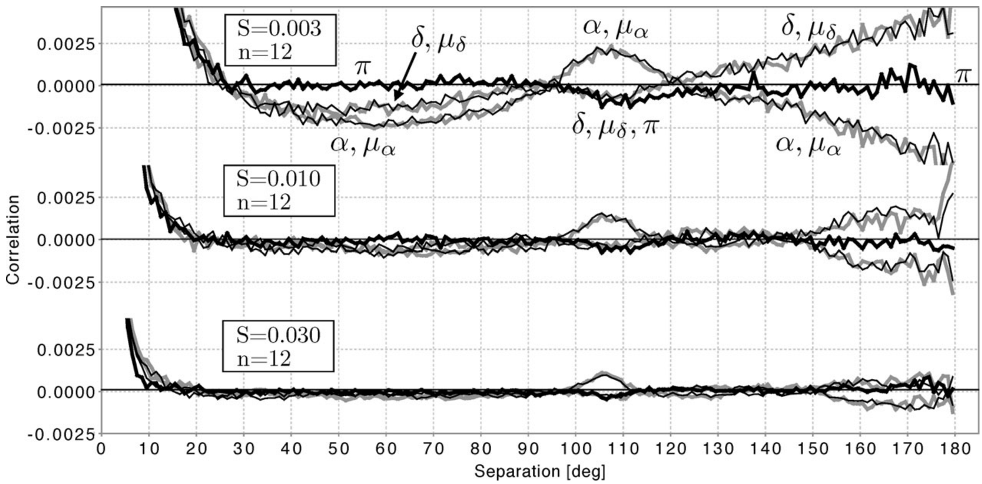

Figure 2. Spatial correlations in experiments A, B and C (see Table 1). Thin black lines are α and δ, the thick black line is π, thick gray lines are μ_α and μ_δ (binsize 1.0°).

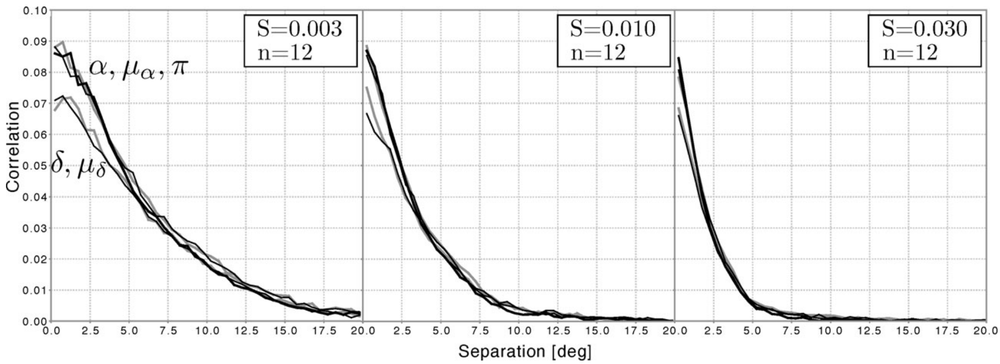

Figure 3. Same as Fig. 2 but plotted for separations $\theta = 0$–20° (binsize 0.5°).

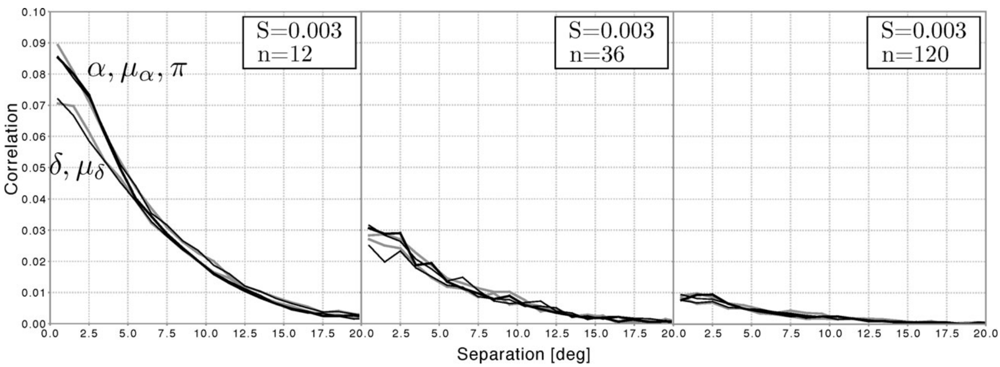

Figure 4. Spatial correlations in experiments C, D and E (see Table 1). Binsize is 1.0°.

References

Hobbs, D., Holl, B., Lindegren, L., Raison, F., Klioner, S., & Butkevich, A. 2009, *this proceedings*, 320

Lindegren, L. 2009, *this proceedings*, 296