

# Joint astrometric solution of HIPPARCOS and *Gaia*

## A recipe for the Hundred Thousand Proper Motions project<sup>★</sup>

Daniel Michalik<sup>1</sup>, Lennart Lindegren<sup>1</sup>, David Hobbs<sup>1</sup>, and Uwe Lammers<sup>2</sup>

<sup>1</sup> Lund Observatory, Lund University, Box 43, 22100 Lund, Sweden  
e-mail: [daniel.michalik;lennart;david]@astro.lu.se

<sup>2</sup> European Space Agency (ESA/ESAC), PO Box 78, 28691 Villanueva de la Cañada, Madrid, Spain  
e-mail: uwe.lammers@sciops.esa.int

Received 15 July 2014 / Accepted 28 July 2014

### ABSTRACT

**Context.** The first release of astrometric data from *Gaia* is expected in 2016. It will contain the mean stellar positions and magnitudes from the first year of observations. For more than 100 000 stars in common with the HIPPARCOS Catalogue it will be possible to compute very accurate proper motions due to the time difference of about 24 years between the two missions. This Hundred Thousand Proper Motions (HTPM) project is planned to be part of the first release.

**Aims.** Our aim is to investigate how early *Gaia* data can be optimally combined with information from the HIPPARCOS Catalogue in order to provide the most accurate and reliable results for HTPM.

**Methods.** The Astrometric Global Iterative Solution (AGIS) was developed to compute the astrometric core solution based on the *Gaia* observations and will be used for all releases of astrometric data from *Gaia*. We adapt AGIS to process HIPPARCOS data in addition to *Gaia* observations, and use simulations to verify and study the joint solution method.

**Results.** For the HTPM stars we predict proper motion accuracies between 14 and 134  $\mu\text{s yr}^{-1}$ , depending on stellar magnitude and amount of *Gaia* data available. Perspective effects will be important for a significant number of HTPM stars, and in order to treat these effects accurately we introduce a formalism called scaled model of kinematics (SMOK). We define a goodness-of-fit statistic which is sensitive to deviations from uniform space motion, caused for example by binaries with periods of 10–50 years.

**Conclusions.** HTPM will significantly improve the proper motions of the HIPPARCOS Catalogue well before highly accurate *Gaia*-only results become available. Also, HTPM will allow us to detect long period binary and exoplanetary candidates which would be impossible to detect from *Gaia* data alone. The full sensitivity will not be reached with the first *Gaia* release but with subsequent data releases. Therefore HTPM should be repeated when more *Gaia* data become available.

**Key words.** astrometry – methods: data analysis – methods: numerical – space vehicles: instruments – proper motions – planets and satellites: detection

## 1. Introduction

Stellar proper motions have traditionally been determined by analysing the differences in position at different epochs, often separated by many decades and obtained using vastly different instruments and methods. In this process, parallaxes (and radial motions, albeit relevant to a much lesser extent) were mostly ignored.

With the advent of space astrometry, most notably the European satellite HIPPARCOS (1989–1993, see [ESA 1997](#)), it became necessary to treat data in a unified manner, i.e., by applying a single least-squares solution for the position, parallax, and annual proper motion. HIPPARCOS determined these parameters for nearly 120 000 stars<sup>1</sup> mostly brighter than magnitude 12, with a median uncertainty of about 1 milli-arcsecond (mas). The *Tycho-2* Catalogue ([Høg et al. 2000](#)) gave additional data for 2.5 million stars observed with the HIPPARCOS starmappers. The re-reduction of the HIPPARCOS raw data ([van Leeuwen 2007a,b](#)) significantly improved the main-mission results. Today, 25 years

after the launch of the satellite, these catalogues remain the main source for the astrometric parameters of these stars.

The European space astrometry mission *Gaia* will soon change this picture. *Gaia*, launched at the end of 2013, will determine the astrometric parameters of up to a billion stars between magnitude 6 and 20 with unprecedented accuracies reaching a few tens of micro-arcseconds ( $\mu\text{s}$ ) for *Gaia* magnitude  $G \lesssim 15$ . The vast amounts of data will be processed in a single coherent least-squares solution, which solves not only for the astrometric parameters but also for a large number of parameters describing the time-varying spacecraft attitude and the geometry of the optical instrument. Due to the very large number of parameters to be determined from the observational data the system cannot be solved directly ([Bombrun et al. 2010](#)) but has to be tackled in a block-iterative manner with the so called “Astrometric Global Iterative Solution” (AGIS). The AGIS software has been designed and implemented by groups at ESA/ESAC, Lund Observatory, and others, and is described in detail together with the fundamental algorithms and mathematical framework by [Lindegren et al. \(2012\)](#).

Astrometric measurements obtained in the past, even of moderate accuracy by modern standards, have lasting value as they represent a state of the Universe that is never repeated. A good example is the construction of proper motions in the

<sup>★</sup> Appendices are available in electronic form at <http://www.aanda.org>

<sup>1</sup> We use “star” to denote a catalogue entry even when it refers to a non-single or extragalactic object. In the context of *Gaia* data processing the term “source” is commonly used for such objects.

*Tycho-2* Catalogue using HIPPARCOS and century-old photographic positions. When the astrometric parameters are propagated over a long time interval, uncertainties in the tangential and radial motions accumulate to a significant positional uncertainty. Yet, long-term deviations from linear space motion (e.g., in long-period binaries) increase even more drastically with time. Such deviations might not be detectable within the time spans of the HIPPARCOS or *Gaia* missions individually, but could be detectable by combining the results of the two. Thus, although HIPPARCOS will soon be superseded by *Gaia* in terms of the expected accuracies at current epochs, its data form a unique comparison point in the past, very valuable in combination with later results. For this reason the first *Gaia* data release scheduled for 2016 will not only publish stellar positions and magnitudes based on the first *Gaia* observations, but also a combination of these observations with the HIPPARCOS Catalogue for all stars common between the two missions. This part of the release is called the Hundred Thousand Proper Motions project (HTPM), originally proposed by F. Mignard in a *Gaia*-internal technical document (Mignard 2009).

The present paper gives a recipe for the practical realisation of the HTPM project in the context of the already existing AGIS scheme for the astrometric solution of *Gaia* data. The proper motions in HTPM might be trivially computed from the positional differences between an early *Gaia* solution and the HIPPARCOS Catalogue – the “conventional catalogue combination” approach of Sect. 2.3. However, we argue that the more elaborate “joint solution” method described in Sect. 2.4 will have important advantages for the HTPM project, and in Sect. 3 we show how to implement it as part of AGIS. The validity and accuracy of the method is demonstrated by means of a joint solution of simulated *Gaia* observations of the HIPPARCOS stars (Sect. 4). In the final sections we discuss the limitations of the results and their validity in the light of *Gaia*’s full nominal mission performance, as well as possible applications of the joint solution method to other astrometric data.

The HTPM project should use the re-reduction of the raw HIPPARCOS data (van Leeuwen 2007b), as it represents a significant improvement over the original HIPPARCOS Catalogue (ESA 1997). Therefore it is also used in all our simulations. For the purpose of demonstrating the HTPM solution we regard all valid entries of the HIPPARCOS Catalogue as astrometrically well-behaved (effectively single) stars. Their space motions are therefore regarded as uniform (rectilinear, with constant speed) over the time interval covered by HIPPARCOS and *Gaia*. This is obviously a very simplified picture of the true content of the HIPPARCOS Catalogue. However, getting the solution right in this simple case is a first necessary step for any more sophisticated treatment of detected binaries and multiple stars in the HIPPARCOS Catalogue.

## 2. Theory

Combining astrometric catalogues requires that data are expressed in the same reference system and described in terms of a common kinematic model. In this section we describe the adopted model and how it is connected to the definition of the astrometric parameters. We outline the conventional approach to catalogue combination and develop the “joint solution” as an optimal generalisation of the method. We show how to detect deviations from the kinematic model or misfits between the datasets. We also outline how to reconstruct the required information from HIPPARCOS and how to integrate the proposed scheme in the astrometric solution algorithm of *Gaia*.

### 2.1. Kinematic model of stellar motion

The choice of astrometric parameters is a direct result of choosing a model of stellar motion. The most basic assumption is for stars to move uniformly, i.e., linearly and with constant speed, relative to the solar system barycentre (SSB). Note that this also means that the stars are assumed to be single. This is obviously not true for all of them, but a good basic assumption for most stars. During the data reduction stars that are not “well behaved” in an astrometric sense can be filtered out and treated further, e.g., by adding additional parameters for components of stellar systems or for acceleration through external influences.

A uniform space motion can be fully described by six parameters: three for the position in space at a chosen reference epoch, and three for the velocity. Traditionally, the three positional parameters are right ascension  $\alpha$ , declination  $\delta$ , and parallax  $\varpi$  relative to the SSB at the reference epoch of the catalogue. The motion is then described by three parameters, where  $\mu_{\alpha^*} = \dot{\alpha} \cos \delta$  and  $\mu_{\delta} = \dot{\delta}$  are the proper motions in right ascension and declination, respectively, and the third parameter  $\mu_r$  is the radial motion component. The radial component is more commonly given as the radial velocity  $v_r$  in  $\text{km s}^{-1}$ , but in an astrometric context it is conveniently expressed as the radial proper motion (equivalent to the relative change in distance over time, or  $-\dot{\varpi}/\varpi$ )

$$\mu_r = v_r \varpi / A, \quad (1)$$

where  $A$  is the astronomical unit expressed in  $\text{km yr s}^{-1}$ . Only the first five parameters are classically considered astrometric parameters. Based on only a few years of observations it is usually not possible to determine the radial component from astrometry with sufficient accuracy (Dravins et al. 1999). Hence the radial component is better determined by other techniques, i.e., from spectroscopy. For *Gaia* the radial component will be significant for many more stars, although the affected fraction remains very small (de Bruijne & Eilers 2012). Even though  $\mu_r$  is not determined in the astrometric solution for the vast majority of sources, it is convenient and sometimes necessary to formulate astrometric problems with the full set of six astrometric parameters, as we do in this paper. We will also show how to treat the sixth component when the radial velocity is unknown or added from spectroscopy.

### 2.2. Dealing with non-linearities: SMOK

When comparing and subsequently combining astrometric catalogues one needs to deal with the fact that the mapping from rectilinear to spherical coordinates is strongly non-linear. This becomes significant at the  $\mu\text{as}$  level when the differences in  $\alpha$  and  $\delta$  exceed some  $(1 \mu\text{as})^{1/2} \approx 0.5 \text{ arcsec}$ . For example, the barycentric direction traced out in  $\alpha(t)$ ,  $\delta(t)$  due to the proper motion will not be linear even though the star is assumed to move uniformly through space. The traditional way to deal with this is to introduce higher-order correction terms computed by Taylor expansion of the rigorous equations (e.g., Taff 1981). In this paper we take a different approach, based on the scaled modelling of kinematics (SMOK) concept described in Appendix A. For the present purpose it is sufficient to know that  $(\alpha, \delta)$  may be replaced by linear coordinates  $(a, d)$  relative to a designated, fixed comparison point, with time derivatives  $\dot{a}$ ,  $\dot{d}$  representing the components of proper motion in  $\alpha$  and  $\delta$ . The six parameters  $a$ ,  $d$ ,  $\varpi$ ,  $\dot{a}$ ,  $\dot{d}$ ,  $\dot{r}$  (where  $\dot{r}$  is the SMOK equivalent of the radial proper motion) provide an alternative and equivalent parametrisation of the kinematics, more convenient for the catalogue combination than the usual set  $\alpha, \delta, \varpi, \mu_{\alpha^*}, \mu_{\delta}, \mu_r$ .

### 2.3. Conventional catalogue combination

In the conventional catalogue combination the astrometric parameters in each catalogue are independently estimated from separate sets of observations, and the combination is done a posteriori from the individual catalogues. Let  $(a_1, d_1, \varpi_1)$  at time  $t_1$  be the position and parallax of a star in the first catalogue, and  $(a_2, d_2, \varpi_2)$  at time  $t_2$  the corresponding information in the second catalogue. The proper motion parameters  $\dot{a}$ ,  $\dot{d}$  are then derived as the positional difference over time  $\Delta t = t_2 - t_1$

$$\dot{a} = (a_2 - a_1)/\Delta t, \quad \dot{d} = (d_2 - d_1)/\Delta t, \quad (2)$$

which is possible thanks to the reformulation of the astrometric parameters in SMOK. The proper motion uncertainties are

$$\sigma_{\dot{a}} = \frac{\sqrt{\sigma_{a1}^2 + \sigma_{a2}^2}}{\Delta t}, \quad \sigma_{\dot{d}} = \frac{\sqrt{\sigma_{d1}^2 + \sigma_{d2}^2}}{\Delta t}, \quad (3)$$

where  $\sigma_{a1}$  is the uncertainty of  $a_1$ , etc. The third kinematic parameter  $\dot{r}$  for the radial motion could in theory be derived from the (negative, relative) difference in parallax, but in practice it is derived from the spectroscopic radial velocity as discussed in Sect. 2.1.

While the proper motions are obtained by taking position differences over time, the combined parameters for position and parallax are formed as weighted means. For  $a$  this gives

$$\hat{a} = \frac{a_1\sigma_{a1}^{-2} + a_2\sigma_{a2}^{-2}}{\sigma_{a1}^{-2} + \sigma_{a2}^{-2}}, \quad (4)$$

referring to the mean epoch of the combination

$$\hat{t}_a = \frac{t_1\sigma_{a1}^{-2} + t_2\sigma_{a2}^{-2}}{\sigma_{a1}^{-2} + \sigma_{a2}^{-2}}. \quad (5)$$

The reference time  $\hat{t}_a$  is the optimal time in-between the two catalogues at which the position and proper motion are uncorrelated and the uncertainty of  $\hat{a}$  is minimal, given by  $\sigma_{\hat{a}}^{-2} = \sigma_{a1}^{-2} + \sigma_{a2}^{-2}$ . The expressions for  $\hat{d}$  and  $\hat{\varpi}$  are analogous.

This combination scheme has some limitations, in that it does not take correlations between the astrometric parameters into account, nor the individual proper motions that may exist in each catalogue. In the next section we describe a more general approach.

### 2.4. Joint solution

The reduction of astrometric data is typically done using least-squares solutions, resulting in a linear system of normal equations  $N\mathbf{x} = \mathbf{b}$ . Here,  $\mathbf{x}$  is the vector of resulting astrometric parameters,  $N$  the normal equations matrix, and  $\mathbf{b}$  a vector constructed from the residuals of the problem<sup>2</sup>. The covariance  $\mathbf{C}$  of the solution  $\hat{\mathbf{x}} = N^{-1}\mathbf{b}$  is formally given by  $\mathbf{C} = N^{-1}$ .

In AGIS the observations of all well-behaved stars (“primary sources”) must be considered together in a single, very large least-squares solution (Sect. 2.7). For  $n$  primary sources,  $\mathbf{x}$  would then be the full vector of  $6n$  astrometric parameters,

with  $N$  and  $\mathbf{b}$  of corresponding dimensions. However, for the present exposition it is sufficient to consider one star at a time, so that  $\mathbf{x}$  and  $\mathbf{b}$  are of length 6 and  $N$  has dimensions  $6 \times 6$ . In practice only five of the six parameters are estimated, and  $N^{-1}$  should hereafter be regarded as the inverse of the upper-left  $5 \times 5$  submatrix<sup>3</sup>.

On the assumption that the adopted kinematic model is valid for a particular star, the matrix  $N$  and vector  $\mathbf{b}$  encapsulate the essential information on the astrometric parameters, as determined by the least-squares solution. Thus, in order to make optimal use of the HIPPARCOS data for a given star there is no need to consider the individual observations of that star: all we need is contained in the “information array”  $[N \mathbf{b}]$ . In Sect. 2.6 we show how this array is reconstructed from the published HIPPARCOS Catalogue.

Let  $[N_1 \mathbf{b}_1]$  and  $[N_2 \mathbf{b}_2]$  be the information arrays for the same star as given by two independent astrometric catalogues. From the way the normal equations are calculated from observational data it is clear that the information arrays are additive, so that  $[N_1 \mathbf{b}_1] + [N_2 \mathbf{b}_2]$  is the information array that would have resulted from processing the two datasets together. In Michalik et al. (2012) we have proposed that the optimum combination of the catalogues is done a priori, that is by adding the corresponding arrays *before* solving. The result,

$$\hat{\mathbf{x}} = (N_1 + N_2)^{-1}(\mathbf{b}_1 + \mathbf{b}_2), \quad (6)$$

is the *joint solution* of the astrometric parameters, with covariance  $\hat{\mathbf{C}} = (N_1 + N_2)^{-1}$ . The two catalogue entries for the star must use the same reference epoch and the same SMOK comparison point.

The joint solution has several advantages over the conventional combination method outlined in Sect. 2.3. Because it uses the full information in each catalogue it makes better use of the data and allows one to estimate the resulting uncertainties more accurately, taking the correlations into account. The individual proper motion information available in each catalogue is automatically incorporated in the joint proper motion. Moreover, a solution might be possible where the data in each set individually is insufficient to solve for all astrometric parameters, that is,  $N_1 + N_2$  may be non-singular even if  $N_1$ ,  $N_2$ , or both, are singular. In practice, if  $N_1$  comes from the HIPPARCOS data, it will always be non-singular (since there is a HIPPARCOS solution), and the sum is then also non-singular. Hence it will always be possible to make a joint solution for all five astrometric parameters of the HTPM stars. Finally, the joint solution scheme is a clean and rigorous approach and can be integrated into the existing implementation of the astrometric solution for *Gaia* with moderate effort.

The joint solution can be seen as a multidimensional generalisation of the conventional scheme in Sect. 2.3, with  $N$  representing the weights ( $\sigma^{-2}$ ) and  $\mathbf{b}$  the astrometric parameters multiplied by their weights (e.g.,  $a\sigma^{-2}$ ). Then Eq. (6) is the matrix-equivalent of Eq. (4). The joint solution can also be understood in terms of Bayesian estimation theory (assuming multivariate Gaussian parameter errors), with  $N_1, \mathbf{b}_1$  representing the prior information,  $N_2, \mathbf{b}_2$  the new data, and their sums the posterior information.

<sup>2</sup> The least squares problem can be solved using a number of alternative numerical algorithms, for example based on orthogonal transformations. However, as these algorithms are all *mathematically* equivalent to the use of normal equations, our results remain valid independent of the chosen solution algorithm.

<sup>3</sup> The full matrix is nevertheless needed for the covariance propagation in Sect. 2.6.

## 2.5. Goodness of fit of the joint solution

The goodness of fit of a least-squares solution can be described in terms of the sum of the squares of the normalized post-fit residuals,

$$Q = \sum_k \left( \frac{\eta_k^{(\text{obs})} - \eta_k^{(\text{calc})}}{\sigma_k} \right)^2, \quad (7)$$

where  $\eta_k^{(\text{obs})}$  and  $\eta_k^{(\text{calc})}$  are the observed and calculated (fitted) angular focal-plane coordinates of the star in observation  $k$ , and  $\sigma_k$  is the standard error of the observation.  $Q$  is calculated for each star separately and is simply a function of  $\mathbf{x} = (a, d, \varpi, \dot{a}, \dot{d}, \dot{r})'$ . The least-squares solution  $\hat{\mathbf{x}} = N^{-1}\mathbf{b}$  minimizes  $Q$  and for any other parameter vector  $\mathbf{x}$  we have

$$Q(\mathbf{x}) = Q(\hat{\mathbf{x}}) + (\mathbf{x} - \hat{\mathbf{x}})'N(\mathbf{x} - \hat{\mathbf{x}}). \quad (8)$$

If the kinematic model is correct and the standard errors of the observations are correctly estimated one expects the minimum value  $Q(\hat{\mathbf{x}})$  to follow the chi-square distribution with  $\nu$  degrees of freedom,  $Q(\hat{\mathbf{x}}) \sim \chi^2(\nu)$ . Here  $\nu = m - \text{rank}(N)$  is equal to the number of observations  $m$  (that is the number of terms in Eq. (7)) diminished by the rank of  $N$ . Note that this holds even if  $N$  is singular (i.e.,  $\text{rank}(N) < n$ , where  $n$  is the number of fitted parameters). In the singular case  $\hat{\mathbf{x}}$  is not unique, yet  $Q(\hat{\mathbf{x}})$  has a well-defined value (which may be 0 or positive).

Analogous to Eq. (8), in the joint solution we minimize the total goodness of fit,

$$Q(\mathbf{x}) = Q_1(\mathbf{x}) + Q_2(\mathbf{x}) = Q_1(\hat{\mathbf{x}}_1) + Q_2(\hat{\mathbf{x}}_2) + (\mathbf{x} - \hat{\mathbf{x}}_1)'N_1(\mathbf{x} - \hat{\mathbf{x}}_1) + (\mathbf{x} - \hat{\mathbf{x}}_2)'N_2(\mathbf{x} - \hat{\mathbf{x}}_2). \quad (9)$$

Here  $\hat{\mathbf{x}}_i = N_i^{-1}\mathbf{b}_i$  is the solution obtained by using only catalogue  $i = 1, 2$ , i.e., minimizing  $Q_i(\mathbf{x})$ , which results in the minimum value  $Q_i(\hat{\mathbf{x}}_i)$ . It is readily seen that Eq. (9) is minimized precisely for the joint solution vector in Eq. (6).

Each of the four terms in Eq. (9) has a simple interpretation. The first term,  $Q_1(\hat{\mathbf{x}}_1)$ , is the chi-square obtained when fitting the astrometric parameters only to the first set of data (in our case the HIPPARCOS data); similarly,  $Q_2(\hat{\mathbf{x}}_2)$  is the chi-square obtained when fitting only to the second set of data (from *Gaia*). The sum of the last two terms is minimized for  $\mathbf{x} = \hat{\mathbf{x}}$ , and shows how much the chi-square is increased by forcing the *same* parameters to fit *both* sets of data in the joint solution. This quantity is useful for assessing whether the two datasets are mutually consistent and we therefore introduce a separate notation for it,

$$\Delta Q = (\hat{\mathbf{x}} - \hat{\mathbf{x}}_1)'N_1(\hat{\mathbf{x}} - \hat{\mathbf{x}}_1) + (\hat{\mathbf{x}} - \hat{\mathbf{x}}_2)'N_2(\hat{\mathbf{x}} - \hat{\mathbf{x}}_2). \quad (10)$$

The two terms give the increase in chi-square due to the first and second dataset, respectively.

Long-period astrometric binaries may have significantly different proper motions at the HIPPARCOS and *Gaia* epochs, and these in turn may differ from the mean proper motion between the epochs. If the differences are significant, compared with the measurement precisions, they will result in an increased value of  $\Delta Q$ . The null hypothesis, namely that the star is astrometrically well-behaved, should be rejected if  $\Delta Q$  exceeds a certain critical value. In order to calculate the critical value it is necessary to know the expected distribution of  $\Delta Q$  under the null hypothesis.

Let  $m_i$  and  $\nu_i = m_i - \text{rank}(N_i)$  be the number of observations and degrees of freedom in catalogue  $i$ . The number of degrees of freedom in the joint solution is  $\nu = (m_1 + m_2) - \text{rank}(N_1 + N_2)$ .

Under the null-hypothesis we have  $Q_i(\hat{\mathbf{x}}_i) \sim \chi^2(\nu_i)$  ( $i = 1, 2$ ),  $Q(\hat{\mathbf{x}}) \sim \chi^2(\nu)$ , and consequently

$$\Delta Q \sim \chi^2(k), \quad (11)$$

where

$$k = \nu - \nu_1 - \nu_2 = \text{rank}(N_1) + \text{rank}(N_2) - \text{rank}(N_1 + N_2). \quad (12)$$

In the special case when  $N_1$ ,  $N_2$ , and  $N_1 + N_2$  all have full rank (equal to  $n$ , the number of astrometric parameters) we have  $k = n$ . At a significance level of 1% the critical values of  $\Delta Q$ , above which the null hypothesis should be rejected, are 15.086, 13.277, 11.345, 9.210, and 6.635 for  $k = 5, 4, 3, 2$ , and 1, respectively (e.g., Abramowitz & Stegun 2012). With this criterion only 1% of the well-behaved stars should be accidentally misclassified as not well-behaved. The expected distribution of  $\Delta Q$  can be verified in the simulations which, by design, only includes well-behaved stars.

## 2.6. Reconstruction of $N_H$ , $\mathbf{b}_H$ for the HIPPARCOS Catalogue

When using the joint solution for incorporating HIPPARCOS data in the solution of early *Gaia* data it is necessary to reconstruct the normal matrix  $N_H$  and the right hand side  $\mathbf{b}_H$  from HIPPARCOS for each star. These are initially calculated for the reference epoch of the HIPPARCOS Catalogue (J1991.25) and later propagated to the adopted reference epoch of the joint solution (see Sect. 2.7).

Let  $a_H, d_H, \varpi_H, \dot{a}_H, \dot{d}_H$  be the astrometric parameters from the HIPPARCOS Catalogue after transformation into the SMOK notation (see Appendix A). The upper-left  $5 \times 5$  submatrix of the covariance matrix can be taken without changes from the HIPPARCOS Catalogue (see Appendix B for details) since  $\sigma_{\alpha^*} = \sigma_a, \sigma_\delta = \sigma_d, \dots$  with sufficient accuracy at the reference epoch of the catalogue and provided that the SMOK comparison point is close enough to the astrometric parameters of the star. The sixth parameter  $\dot{r}_H$  and its corresponding entries in the covariance matrix need to be added from external sources or set to sensible values if not available (see below). Then the normal matrix is simply the inverse of the covariance matrix  $N_H = C_H^{-1}$  and

$$\mathbf{b}_H = N_H \left( a_H, d_H, \varpi_H, \dot{a}_H, \dot{d}_H, \dot{r}_H \right)'. \quad (13)$$

ESA (1997), Vol. 1, Eq. (1.5.69) shows how to reconstruct the elements  $[C_0]_{i6} = [C_0]_{6i}$  ( $i = 1 \dots 6$ ), that is the sixth column and row of the covariance matrix corresponding to the radial motion  $\mu_r$ . Let  $\bar{v}_r, \bar{\varpi}, \bar{\mu}_r$  be the true values and  $\delta v_r, \delta \varpi, \delta \mu_r$  the errors. The expression in Eq. (1.5.69) for the diagonal element  $[C_0]_{66}$  is only valid if the relative uncertainties in the radial velocity and parallax are small, i.e.,  $|\delta v_r/\bar{v}_r|, |\delta \varpi/\bar{\varpi}| \ll 1$ . If this is not the case we need to consider the complete expression for the calculated radial motion,

$$\mu_r = \bar{\mu}_r + \delta \mu_r = (\bar{v}_r + \delta v_r)(\bar{\varpi} + \delta \varpi)/A, \quad (14)$$

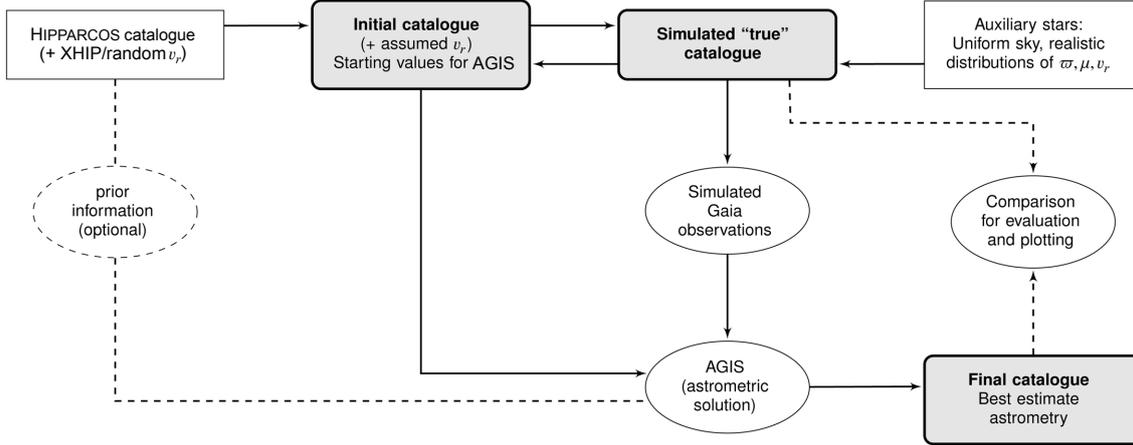
where  $\bar{\mu}_r = \bar{v}_r \bar{\varpi}/A$ , leading to

$$\delta \mu_r = (\bar{v}_r \delta \varpi + \bar{\varpi} \delta v_r + \delta v_r \delta \varpi)/A. \quad (15)$$

Squaring and taking the expectation while assuming that the errors in parallax and radial velocity are uncorrelated gives

$$E(\delta \mu_r^2) = \left( E(v_r^2 \delta \varpi^2) + E(\varpi^2 \delta v_r^2) + E(\delta v_r^2 \delta \varpi^2) \right) / A^2, \quad (16)$$

where we replaced the true quantities by the observed ones. The third term is the required generalisation if  $v_r$  or  $\varpi$  is zero, or if



**Fig. 1.** Relationships between catalogues during simulation runs.

the relative errors are large. For example, if parallax and radial motion are unknown they could be assumed to be zero with a large uncertainty. The generalized version of Eq. (1.5.69) in [ESA \(1997\)](#) reads

$$\begin{aligned} [C_0]_{66} &= (v_{r0}/A)^2 [C_0]_{33} + (\varpi_0/A)^2 \sigma_{vr0}^2 + (\sigma_{vr0}/A)^2 [C_0]_{33}, \\ [C_0]_{i6} &= [C_0]_{6i} = (v_{r0}/A) [C_0]_{i3}, \quad i = 1 \dots 5. \end{aligned} \quad (17)$$

The HIPPARCOS Catalogue contains numerous entries for non-single stars, for which additional parameters are given, describing deviations from uniform space motion. These additional parameters are ignored in our simulations, which regard every star as single. In the actual HTPM solution many of these stars may require more specialised off-line treatment. This is not further discussed in this paper.

### 2.7. Joint solution in AGIS

In reality the astrometric solution cannot be done separately for each star as described in Sect. 2.4 but must consider all the stars together with the spacecraft attitude and instrument calibration. Without prior information on the astrometric parameters this leaves the solution undetermined with respect to the reference frame. This is not the case for the joint solution, however, as the HIPPARCOS prior information contains positions and proper motions that are expressed in a specific reference frame, namely the HIPPARCOS realisation of the International Celestial Reference System (ICRS; [Feissel & Mignard 1998](#)). The incorporation of the HIPPARCOS prior in the joint solution automatically ensures that the resulting data are on the HIPPARCOS reference frame. If required, the data can later be transformed into a more accurate representation of the ICRS (see Sect. 5.3).

Due to the size of the data reduction problem, AGIS does not directly solve  $N\mathbf{x} = \mathbf{b}$  but iteratively improves the astrometric parameters by computing the updates  $\Delta\mathbf{x}$ , i.e., the difference to the current best estimate values. When incorporating HIPPARCOS data this requires us to also express the HIPPARCOS data (subscript H) as a difference to the current best estimate (subscript c). Therefore we construct

$$\Delta\mathbf{b}_H = N_H \Delta\mathbf{x} = N_H \begin{pmatrix} a_H - a_c \\ d_H - d_c \\ \varpi_H - \varpi_c \\ \dot{a}_H - \dot{a}_c \\ \dot{d}_H - \dot{d}_c \\ \dot{r}_H - \dot{r}_c \end{pmatrix}. \quad (18)$$

Before solving we add the corresponding matrices for the *Gaia* data. If no additional *Gaia* data would be added the solution would immediately recover the HIPPARCOS Catalogue parameters.

The reference epoch of the joint solution can be arbitrarily chosen. In practice the *Gaia* data are much better than the HIPPARCOS data, therefore the optimal reference epoch would always be very close to the epoch of the *Gaia* data alone. Assuming one releases *Gaia*-only data and HTPM results at the same time it might be convenient to publish both for the same reference epoch, i.e., the *Gaia*-only reference epoch of the data release.

## 3. Simulations

### 3.1. Logic of simulations

Simulations are based on AGISLab ([Holl et al. 2012](#)), a small-scale version of the AGIS data reduction created and maintained at Lund Observatory. It is used to aid the development of algorithms for the astrometric data reduction of *Gaia*. Simulation runs are carried out in the following steps (cf. Fig. 1):

1. Creating catalogues of all the stars used in the simulation, namely the HIPPARCOS stars and the auxiliary stars (see below). Two catalogues are needed: a simulated “true” catalogue to generate *Gaia* observations and to evaluate the uncertainties of the astrometric performance, and an initial catalogue of starting values for the data reduction.
2. Simulating observations of the stars using the Nominal Scanning Law ([de Bruijne et al. 2010](#)), including perturbations according to the expected precision of *Gaia* measurements.
3. Improving the astrometry of the initial catalogue through the astrometric solution (AGIS), resulting in the final catalogue. This can be done with or without incorporation of prior information from HIPPARCOS.
4. Evaluating the error of the resulting solution by comparing the final catalogue with the true catalogue.

Details of the first two steps are given below, while remaining steps are covered in Sect. 4.

### 3.2. Simulating the stellar catalogues

All catalogues consist of two parts, the HIPPARCOS stars and the additional auxiliary stars. The HIPPARCOS stars are necessary for the realisation of the HTPM scheme, and 113 396 stars

are within the nominal magnitude range of *Gaia* ( $G \simeq 6\text{--}20$ ). In order to obtain a reliable astrometric solution with a realistic modelling of the attitude constraints we find that a minimum of one million stars is needed, uniformly distributed on the sky. 886 604 auxiliary stars are therefore added to the HIPPARCOS stars in the solution. The astrometric results for the auxiliary stars are not included in the statistics for the HTPM performance, which is based only on the results for the HIPPARCOS stars. However, they contribute indirectly to the HTPM solution via the attitude.

### 3.2.1. Simulated “true” catalogue

The true catalogue defines the stars used for creating the simulated *Gaia* observations. For the real mission the true catalogue is of course not known.

To derive the HIPPARCOS portion of the true catalogue we assume that the true parameters deviate from the HIPPARCOS values by random amounts consistent with the HIPPARCOS covariances. The HIPPARCOS Catalogue is taken from CDS and contains the astrometric parameters for the reference epoch J1991.25, including their covariance matrices (Appendix B). For each star let  $\mathbf{C}$  be its covariance matrix,  $\mathbf{L}$  the lower triangular matrix resulting from the Cholesky decomposition  $\mathbf{C} = \mathbf{L}\mathbf{L}'$ , and  $\mathbf{g}$  a vector of six independent standard Gaussian random variables (zero mean, unit standard deviation). Then the true parameters (subscript T) are obtained by applying the error vector  $\mathbf{e} = \mathbf{L}\mathbf{g}$  to the astrometric parameters from the HIPPARCOS Catalogue (subscript H):

$$\mathbf{x}_T = \mathbf{x}_H + \mathbf{e}. \quad (19)$$

Since  $E(\mathbf{g}) = \mathbf{0}$ , where  $E(\dots)$  denotes the expectation value, it follows that  $E(\mathbf{e}) = \mathbf{0}$ . Moreover, since  $E(\mathbf{g}\mathbf{g}') = \mathbf{I}$  (the identity matrix), it is readily verified that  $\mathbf{e}$  has the desired covariance  $E(\mathbf{e}\mathbf{e}') = \mathbf{C}$ . For a joint solution with simulated *Gaia* data the HIPPARCOS Catalogue needs to be propagated to the reference epoch used in the solution.

Rigorous propagation of the astrometric parameters must take into account the radial motions of the stars, for which radial velocities are needed. We use data from XHIP (Anderson & Francis 2012), a compilation of radial velocities and other data for the HIPPARCOS stars from 47 different sources. We only use radial velocities with quality flag “A” or “B” in XHIP. This makes for a total of 40 171 radial velocities which are used as true values in our simulations. For the remaining HIPPARCOS stars we assign random radial velocities from a Gaussian distribution with  $v_r = 0$ ,  $\sigma_{v_r} = 30 \text{ km s}^{-1}$  using Eq. (19), based on the assumption that radial velocities are typically smaller than that. The radial velocity uncertainty (taken from XHIP or using  $30 \text{ km s}^{-1}$ ) is also used to expand the  $5 \times 5$  covariance matrix by a sixth column and row for the uncertainty and correlation of the radial motion, using Eq. (17).

For the auxiliary stars, the positions are chosen to give a random uniform distribution across the sky with a mean density of about  $21 \text{ stars deg}^{-2}$ , corresponding to one million stars needed for the solution. We assume magnitude  $G = 13$  for all auxiliary stars. Since the number density of actual stars with  $G \leq 13$  is about  $60 \text{ deg}^{-2}$  at the Galactic poles, the assumed distribution is a rather conservative estimate of the density of bright stars available for the astrometric solution. The parallaxes of the auxiliary stars are assumed to have a log-normal distribution with median parallax  $2.5 \text{ mas}$  and a standard deviation of  $0.6 \text{ dex}^4$ . The true

<sup>4</sup> Neglecting extinction, this corresponds to a Gaussian distribution of absolute magnitudes  $M_G$  with mean value  $+5$  and standard

proper motions and radial velocities are calculated by assuming an isotropic velocity distribution relative to the Sun with a standard deviation of  $30 \text{ km s}^{-1}$ .

### 3.2.2. Initial catalogue and astrometric solution

The initial catalogue contains the starting values for the data processing. The HIPPARCOS portion of it is identical to the astrometric parameters read from the HIPPARCOS Catalogue. For the auxiliary stars the initial positions are obtained by perturbing the true positions with Gaussian noise of standard deviation  $100 \text{ mas}$  in each coordinate, while the initial parallax and proper motion are set to zero. This is similar to a real life scenario where one would assume initial stellar positions from ground based observations or the first published *Gaia* positions without additional knowledge on the parallax or proper motion. The astrometric values in the initial catalogue are subsequently updated by the AGIS processing, resulting in the final catalogue once the solution is found. We do not solve for the radial motion but set the radial velocity to either zero (assuming no knowledge about it) or the true value (assuming it is perfectly known). In the first case perspective acceleration may show up for some stars as discrepancies in the solution, which disappear when the true radial velocities are used instead (see Sect. 4.3).

### 3.2.3. Final catalogue

The final catalogue contains the astrometric parameters after data processing. The difference to the simulated true catalogue gives the final errors of the reduced data and is used to evaluate the quality of the astrometric results. In this evaluation we focus on the improvement in the astrometric parameters of the HIPPARCOS stars.

## 3.3. Simulating *Gaia* observations

The observations of the one million stars described above are simulated using the Nominal Scanning Law of *Gaia*. We neglect so called “dead time” (when no data can be accumulated for example due to orbit maintenance manoeuvres and micro-meteoroid hits), which may amount to up to 15% of the mission time. We do, however, account for the dead time originating from stellar transits coinciding with gaps between the CCDs in the focal plane, i.e., our simulations remove such observations before further processing of the data.

To account for observation noise, i.e., the expected centroiding performance of *Gaia*, we use a simplified noise model that ignores the gating scheme that *Gaia* exploits for bright star detection. This noise model assumes a constant centroiding performance for all HIPPARCOS stars, identical to the centroiding performance for the brightest ungated stars at magnitude 13. The typical along-scan standard error due to photon statistics is  $94 \mu\text{as}$ . A second noise component is added to account for various effects, such as attitude modelling errors (Risquez et al. 2013) and uncertainties originating from geometrical calibration parameters of the spacecraft. Although this additional noise component may be correlated between individual CCD observations, we model it by quadratically adding a conservative RMS value of  $300 \mu\text{as}$  to the photon statistical standard error per CCD.

deviation 3 mag. This is not unreasonable for a local magnitude-limited stellar sample; cf. the HR diagram for nearby HIPPARCOS stars, such as Fig. 1 in Dehnen & Binney (1998). The assumed distribution of true parallaxes and proper motions has some impact on our case B simulation results as discussed in Sect. 5.2.

**Table 1.** Number of astrometric parameters per star estimated in the four astrometric solution scenarios.

	Case A (optimistic)		Case B (conservative)	
	<i>Gaia</i> 12	HTPM	<i>Gaia</i> 12	HTPM
HIPPARCOS stars	5	5	2	5
Auxiliary stars	5	5	2	2

Based on the current *Gaia* data release scenario<sup>5</sup> we assume that the HTPM project will initially be based on one year of *Gaia* data. The simulation results presented in Sect. 4 use one year of *Gaia* observations centred around the adopted reference epoch J2015.0.

## 4. Results

### 4.1. Astrometric solution scenarios

Table 1 gives an overview of the four different solution scenarios investigated in this paper. The two cases called *Gaia* 12 do not use any prior data from the HIPPARCOS Catalogue, but only the 12 months of *Gaia* observations. The other two, called HTPM, use the HIPPARCOS covariances and astrometric parameters as priors in the processing of the same *Gaia* observations as in *Gaia* 12. A comparison between the HTPM and *Gaia* 12 scenarios thus allows one to assess the improvement brought by the HIPPARCOS prior information.

The scenarios are subdivided into cases A and B. In case A we assume that there is sufficient *Gaia* data to perform a full five-parameter astrometric solution for all stars even without the HIPPARCOS prior. This is an optimistic assumption, since in reality one year of data is only barely sufficient for a five-parameter solution under ideal conditions, i.e., without data gaps. Dead time as outlined before and the actual temporal distribution of observations over the year could mean that the solution must be constrained to estimate only the two positional parameters for most of the stars. We simulate this in case B by conservatively assuming that all stars for which we do not include a prior will have a two-parameter solution. In such a solution the parallaxes and proper motions are effectively assumed to be zero, which gives a large additional error component in the estimated positions<sup>6</sup>. While the *Gaia* 12-B solution is then restricted to two parameters for all stars, HTPM-B can still solve all five parameters of the HIPPARCOS stars. Case B might be closer to the foreseen first release of *Gaia* data and the first release of HTPM. Case A on the other hand demonstrates the capabilities of *Gaia* and HTPM once sufficient data for a full astrometric solution are available in subsequent releases of *Gaia* data.

### 4.2. Predicted astrometric accuracies of HTPM

Table 2 summarizes the results for the entire set of HIPPARCOS stars, and subdivided by magnitude. No results are given for

<sup>5</sup> See <http://www.cosmos.esa.int/web/gaia/release> (2014 July 23). The first release of *Gaia* data is foreseen for summer 2016. Discounting in-orbit commissioning, ecliptic pole scanning, and time for data processing leaves us with about one year of *Gaia* data.

<sup>6</sup> Forcing a two-parameter solution in case B for the stars without a prior creates residuals that are much larger than the formal uncertainties of the *Gaia* observations. The astrometric solution copes with this situation by means of the excess noise estimation described in Sect. 3.6 of Lindegren et al. (2012). Effectively this reduces the weight of the *Gaia* observations but does not affect the HIPPARCOS prior. Without excess noise estimation the errors of the HTPM proper motions in case B would be several times larger.

the auxiliary stars, but they are similar to the results for the HIPPARCOS stars in the *Gaia* 12 scenarios. For comparison we also give the formal uncertainties from the HIPPARCOS Catalogue. For the positions they are given both at the original epoch J1991.25 and at the epoch J2015 of the *Gaia* data. It should be noted that the simulations include stars which in the HIPPARCOS Catalogue are described with more than five parameters, but are here treated as single stars. Excluding them from the statistics would systematically reduce the HIPPARCOS uncertainties in Table 2. The real HTPM solution will also include all HIPPARCOS stars independent of the type of solution in the HIPPARCOS Catalogue. A poor fit between the *Gaia* and HIPPARCOS data will then be used to filter out binary candidates for further treatment.

All *Gaia* 12 and HTPM uncertainties in Table 2 are derived from the distribution of the actual errors (calculated values minus true values) obtained in the solutions, using the robust scatter estimate (RSE)<sup>7</sup>. Rather than stating the uncertainty of  $\alpha$  and  $\delta$  separately we give the mean of the RSE in the two coordinates as the position uncertainty. Similarly the proper motion uncertainty is the mean RSE of the errors in  $\mu_{\alpha^*}$  and  $\mu_{\delta}$ .

**Proper motion.** The joint solution shows a big improvement in the proper motion uncertainties compared with the HIPPARCOS data. The improvement factor of HTPM compared with HIPPARCOS alone is 32 in case A and 25 in case B. The factors are similar because the HIPPARCOS position uncertainty dominates over the *Gaia* uncertainty in both cases. In the optimistic case A, the proper motions from the *Gaia*-only data are already better than HIPPARCOS alone, but not as good as the joint HTPM solution.

Using Eq. (3) to estimate the expected precision of the conventional combination we find in case A proper motions of 16 and 137  $\mu\text{as yr}^{-1}$  for the brightest and faintest magnitude bins, compared with 14 and 94  $\mu\text{as yr}^{-1}$  in the HTPM-A results. In case B we find 143 and 602  $\mu\text{as yr}^{-1}$ , respectively, compared with 27 and 134  $\mu\text{as yr}^{-1}$  in HTPM-B. The joint solution thus gives consistently better results, as discussed in Sect. 2.4.

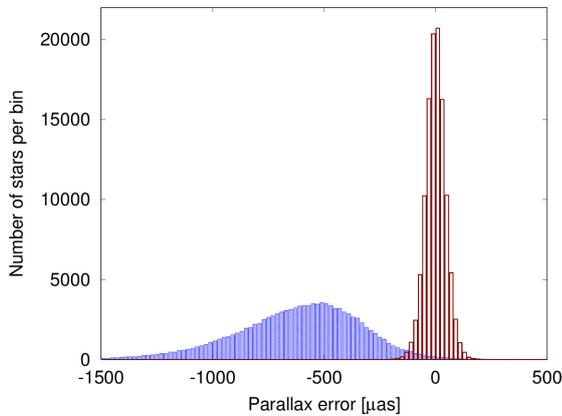
**Parallax.** The improved proper motions allow better to disentangle the five parameters in the joint astrometric solution (cf. Fig. 3), resulting in improved parallax uncertainties. In case A we find that the parallax uncertainties in the joint solution improve by a factor 23 compared with HIPPARCOS, and a factor 2 compared with *Gaia* 12. However, in the more realistic case B the improvement is much smaller (a factor 3 compared with HIPPARCOS) and the parallaxes are strongly biased as shown in Fig. 2. This bias originates from the assumption of zero parallax and proper motion in the two-parameter solution of the auxiliary stars. The true positive parallaxes result in a biased attitude, which propagates into the five-parameter solution of the HIPPARCOS stars making their parallaxes systematically too small. (As discussed in Sect. 5.2, this bias can be entirely avoided in later releases of *Gaia* data through a proper selection of primary sources.)

<sup>7</sup> The RSE is defined as 0.390152 times the difference between the 90th and 10th percentiles of the distribution of the variable. For a Gaussian distribution it equals the standard deviation. Within the *Gaia* core processing community the RSE is used as a standardized, robust measure of dispersion (Lindegren et al. 2012).

**Table 2.** Predicted uncertainties of the astrometric parameters of the HIPPARCOS stars.

Mag.	Number	Position [ $\mu\text{as}$ ]						Parallax [ $\mu\text{as}$ ]				Proper motion [ $\mu\text{as yr}^{-1}$ ]					
		Hip	Hip <sub>2015</sub>	Gaia 12		HTPM		Hip	Gaia 12		HTPM		Hip	Gaia 12		HTPM	
				A	B	A	B		A	B	A	B		A	B		
6–7	9381	367	10 892	41	3388	36	312	501	82	–	43	250 <sup>a</sup>	458	207	–	14	27
7–8	23 679	497	14 434	41	2692	35	318	684	81	–	43	261 <sup>a</sup>	608	204	–	19	30
8–9	40 729	682	19 947	41	2369	35	330	939	77	–	43	271 <sup>a</sup>	840	197	–	26	35
9–10	27 913	936	27 629	40	2663	35	333	1284	77	–	43	274 <sup>a</sup>	1165	194	–	35	43
10–11	8563	1403	41 352	42	5240	36	343	1921	83	–	46	283 <sup>a</sup>	1744	205	–	50	60
11–12	2501	2125	61 896	41	13 687	35	357	2882	78	–	47	291 <sup>a</sup>	2607	195	–	70	85
$\geq 12$	630	3248	109 030	42	13 926	38	378	4291	80	–	51	295 <sup>a</sup>	4578	209	–	94	134
all	113 396	753	22 148	41	2856	35	328	1033	79	–	44	271 <sup>a</sup>	932	199	–	29	38

**Notes.** We compare the HIPPARCOS data alone (Hip) with a solution using only 12 months of *Gaia* data (*Gaia* 12), and a joint solution of HIPPARCOS and *Gaia* data (HTPM). Case A and B refer to the optimistic and conservative scenarios, respectively, described in the text. The two rightmost columns give the predicted HTPM proper motion uncertainties in the two cases. <sup>(a)</sup> Case B parallaxes are biased as shown in Fig. 2. This bias is not included in the RSE values given here.



**Fig. 2.** Parallax errors in the HTPM solution for two cases. Bin width is  $20 \mu\text{as}$ . In case A (full five-parameter astrometric solution for all stars, red/right histogram) the parallax errors are unbiased. In case B (two-parameter solution of the auxiliary stars, blue/left histogram) the median parallax error is  $-591 \mu\text{as}$ .

**Position.** The extremely good *Gaia* observations lead to an improvement by up to a factor  $\sim 600$  compared with HIPPARCOS positions propagated to J2015. In case A the slight improvement in the HTPM positions compared with *Gaia* 12 comes from the better determination of proper motion and parallax. In case B the *Gaia*-only positions show a high uncertainty due to the two-parameter solution which neglects the true parallaxes and proper motions of the stars. The increase in position uncertainties is especially pronounced for the fainter stars due to preferential selection of nearby high-proper motion stars in the non-survey part of the HIPPARCOS Catalogue, which means that their (neglected) parallaxes and proper motions are statistically much larger than for the brighter (survey) stars. In the HTPM solution for case B all five parameters are solved for the HIPPARCOS stars, so the sizes of their parallaxes and proper motions have no direct impact on the accuracy of the solution. However, the positional uncertainties are still much increased compared with case A, because the two-parameter solutions for the auxiliary stars degrade the attitude estimate.

#### 4.3. Goodness of fit statistics

As discussed in Sect. 2.5 the goodness of fit value  $\Delta Q$  from Eq. (10) describes how well the joint astrometric solution fits

the individual observations of both missions together. If all the observations are consistent with the kinematic model, then  $\Delta Q$  is expected to follow a  $\chi^2$  distribution with five degrees of freedom. Larger values indicate deviations from the model, for example non-uniform motion caused by invisible companions or astrometric binaries. In the present simulations we do not include any such objects, so we expect  $\Delta Q$  to follow the theoretical distribution.

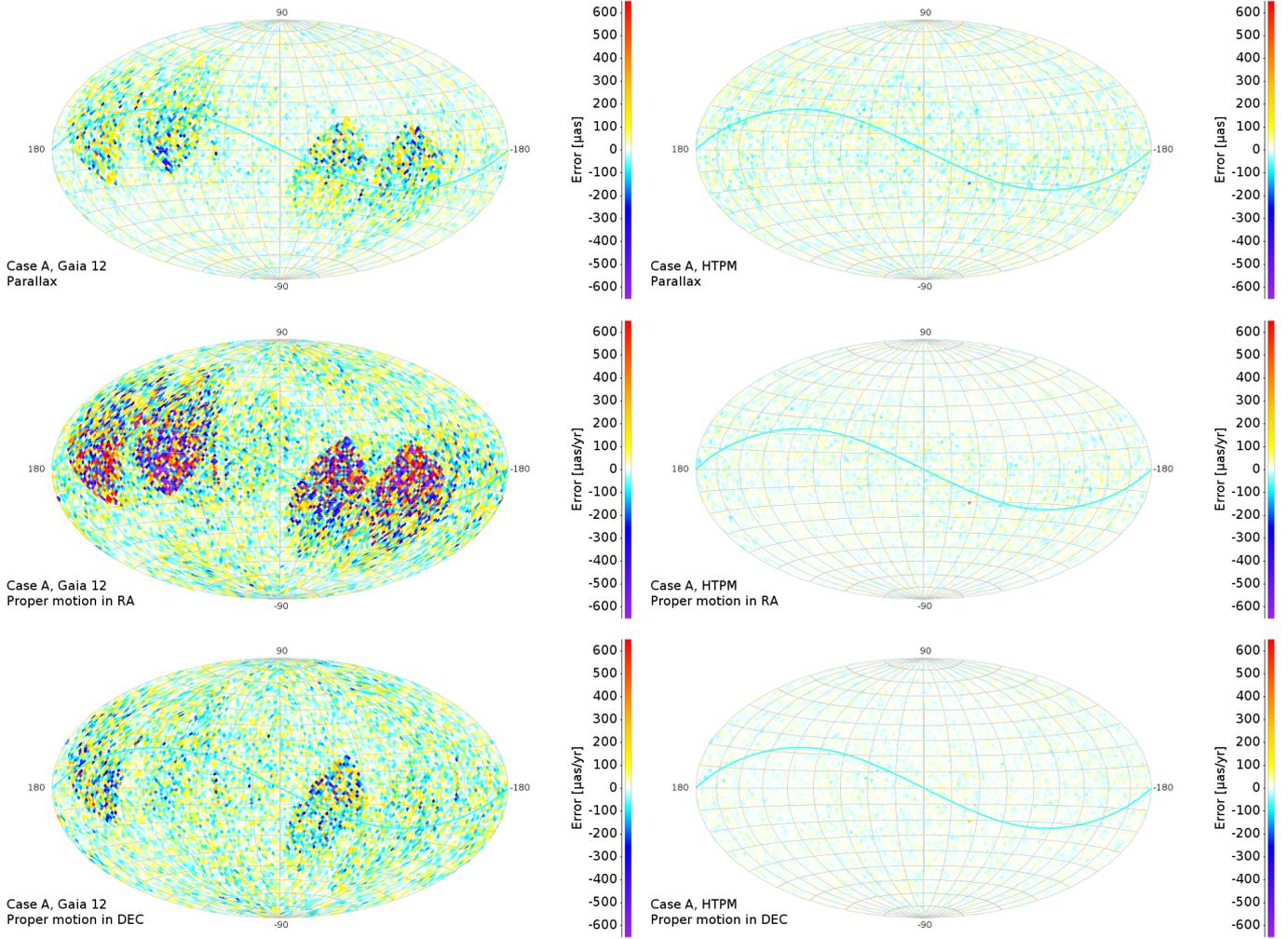
The top two diagrams in the left column in Fig. 4 shows that this is indeed true in case A, if the radial velocities assumed in the solution are the true ones. The result would have been the same if the assumed radial velocities had only been wrong by a few  $\text{km s}^{-1}$ . If instead we assume zero radial velocities for all stars, as was done in the bottom two diagrams (while the observations were still generated with non-zero radial velocities), we find a small number of outliers. It turns out that all of them are nearby, high-velocity stars (Table 3) expected to show significant perspective acceleration, that is the change in proper motion due to the changing stellar distance and the changing angle between the line of sight and motion of the star (Schlesinger 1917; van de Kamp 1977; Murray 1983). This perspective acceleration is not taken into account in the solution when the radial velocities are assumed to be zero, giving a mismatch between the HIPPARCOS data and the observed *Gaia* position. The positional offset due to the perspective acceleration after  $\Delta t$  years amounts to

$$\Delta\theta_{\text{persp}} = \mu\mu_r\Delta t^2, \quad (20)$$

where  $\mu = (\mu_{\alpha^*}^2 + \mu_{\delta}^2)^{1/2}$  is the total proper motion. As shown in Table 3, the stars with a large  $\Delta Q$  also have a large offset  $\Delta\theta_{\text{persp}}$  at the *Gaia* epoch, compared with the positional uncertainty of the solution at that epoch.

This demonstrates that knowledge of radial velocities is required for a number of stars to avoid false positives in the detection of non-uniform space motion (de Bruijne & Eilers 2012). It also shows that  $\Delta Q$  is a useful statistic for detecting non-uniform space motion in general.

The right column in Fig. 4 shows the corresponding results in case B. Here  $\Delta Q$  follows a scaled version of the expected distribution with a somewhat extended tail. The two bottom panels show that  $\Delta Q$  is still a useful measure of deviations from the adopted kinematic model although it is much less sensitive than in case A. As a result only two outliers due to the perspective acceleration are found if the assumed radial velocities are set to zero. This demonstrates the strong dependency of  $\Delta Q$  on the quality of the *Gaia* solution.



**Fig. 3.** Distribution of the parallax and proper motion errors on a Hammer-Aitoff equatorial projection of the sky. All maps are for case A (full five-parameter solutions for all stars). *Left figures:* results from the 12 months’ *Gaia*-only simulation. Some regions of the sky are poorly observed resulting in zonal errors. *Right figures:* HTPM results for the same stars. The prior helps to disentangle proper motion and parallax, therefore we find a more homogeneous distribution of errors at an overall lower level. The cyan line follows the ecliptic for reference.

## 5. Discussion

### 5.1. Longevity of the HTPM solution: detection of binary and exoplanetary candidates

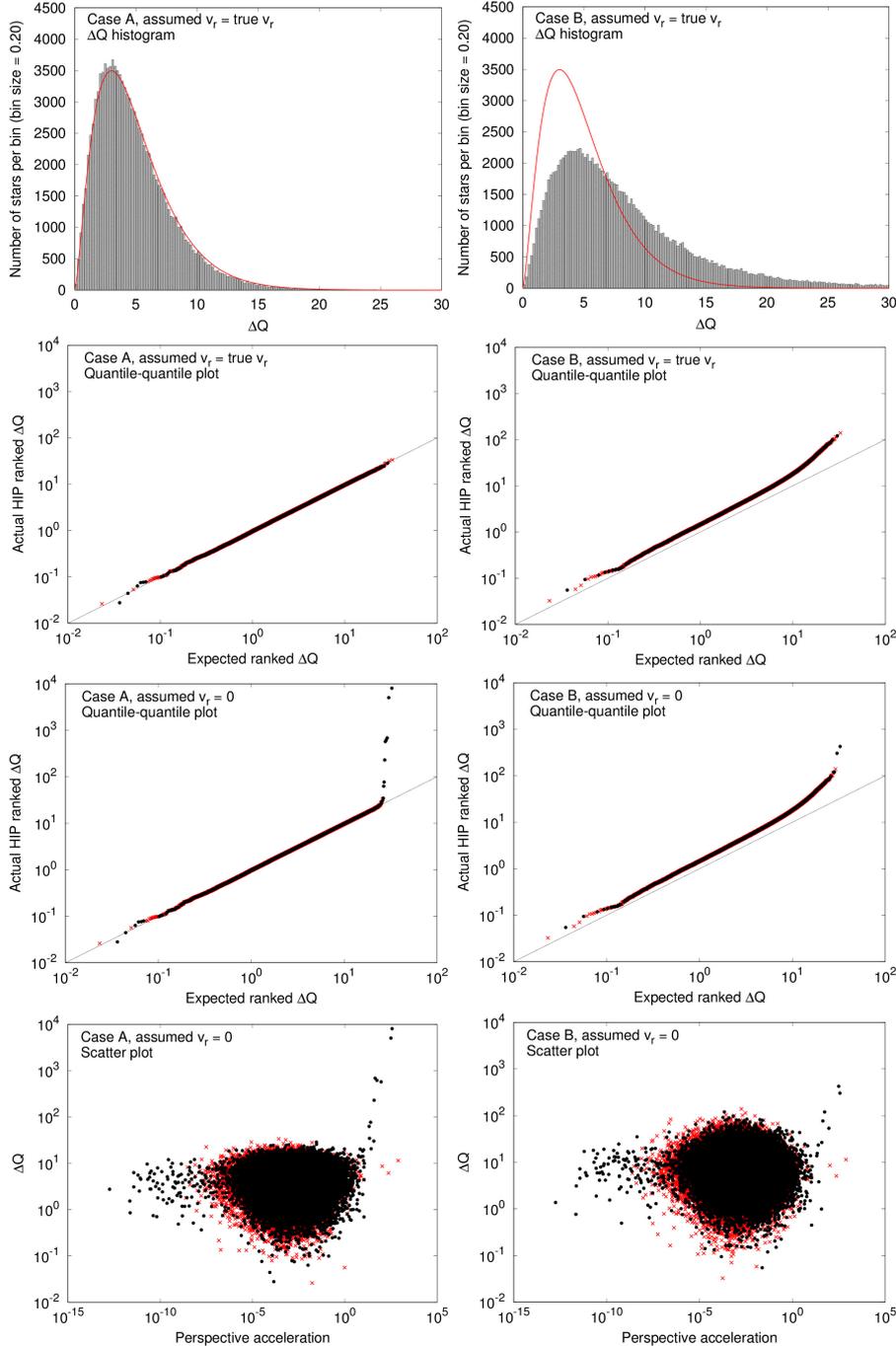
As *Gaia* collects further data the accuracy of the proper motions determined from *Gaia* data alone will eventually supersede that of HTPM. Assuming nominal mission performance and that the proper motion uncertainty scales with mission length as  $L^{-1.5}$ , this will happen already after 2–3 years of *Gaia* data have been accumulated. Still, HTPM will remain a valuable source of information as it is based on a much longer time baseline. This is relevant for long period companions which create astrometric signatures that cannot be seen in *Gaia* data alone. We therefore suggest that HTPM should be repeated with future *Gaia* releases. The goodness-of-fit of the combined solution is sensitive to small deviations of the stellar motions from the assumed (rectilinear) model. This sensitivity will dramatically increase with more *Gaia* data, namely when the *Gaia*-only proper motions become as good as the combined HTPM proper motions.

The potential for detecting faint (stellar or planetary) companions to nearby stars can be illustrated by a numerical example. Consider a  $1 M_{\odot}$  star at 10 pc distance ( $\varpi = 100$  mas) from

the Sun, with an invisible companion of mass  $m$  orbiting at a period of  $P \approx 25$  years (semi-major axis  $a \approx 8.5$  au). The astrometric signature of the companion (i.e., the angular size of the star’s orbit around their common centre of mass; Perryman 2014) is  $a_* \approx a\varpi(m/M_{\odot}) \approx 850(m/M_{\odot})$  mas if the orbit is seen face-on, and the instantaneous proper motion of the star relative to the centre of mass is  $2\pi a_*/P \approx 200(m/M_{\odot})$  mas yr $^{-1}$ . If HIPPARCOS effectively measures this instantaneous proper motion which is extrapolated over  $\Delta t = 25$  years, the extrapolated position from HIPPARCOS (with its uncertainty of about 22 mas, see Table 2) and the position observed by *Gaia* (with an uncertainty much lower than from HIPPARCOS) could differ by up to  $\approx 5000(m/M_{\odot})$  mas. Assuming that detection is possible if the position difference is at least twice as large as the positional uncertainty<sup>8</sup>, we find that the initial HTPM results could be sensitive to companion masses down to  $\approx 10^{-2} M_{\odot}$ , that is brown dwarf or super-Jupiter companions.

If we instead let *Gaia* measure the instantaneous proper motion of the system and propagate backwards to the HIPPARCOS epoch, we can take advantage of the much better uncertainties

<sup>8</sup> Table 3 shows that  $\Delta Q$  in case A may be sensitive to positional deviations at the *Gaia* epoch as small as 21 mas.



**Fig. 4.** *Left column:* Goodness of fit values  $\Delta Q$  for case A simulations. *From top to bottom*, the  $\Delta Q$  values (grey bars) follow a  $\chi^2$  distribution (red line) with five degrees of freedom. If the assumed radial velocities in the solution equal the true values, the actual and expected distributions agree perfectly. If the assumed radial velocity is unknown (set to zero) deviations from the expected distributions are seen. These outliers are caused by perspective acceleration. The markers in the quantile-quantile and scatter plots correspond to stars with radial velocities from XHIP (black dots) and to stars with random radial velocities (red crosses). The three rightmost red crosses in the scatter plots correspond to HIP 80190, HIP 80194, and HIP 67694 which have very large uncertainties in the HIPPARCOS Catalogue. Therefore they do not show a large  $\Delta Q$  value even though they have large perspective acceleration. *Right column:* same plots for case B simulations (see Sect. 2.5).

of the *Gaia* astrometry. Two to three years of *Gaia* data already give proper motion uncertainties better than  $30 \mu\text{s yr}^{-1}$  for the bright stars, and hence extrapolated position uncertainties better than HIPPARCOS at its own epoch, or  $\approx 0.75 \text{ mas}$  (Table 2). Therefore the HTPM sensitivity increases roughly by a factor 30, allowing the detection of companion masses down to about  $3 \times 10^{-4} M_{\odot}$ , or Saturn-type objects at a Saturn-like distance to the host star.

This demonstrates that the results of HTPM can be used to find candidates for long period exoplanets around nearby stars, with a highly interesting companion mass range opening up with subsequent releases of *Gaia* data when combined with HIPPARCOS. These companions cannot be detected from *Gaia* data alone even at the full mission length, and are hard to detect through classical methods due to their long periods, low transit probability and small radial velocity signatures. Since  $\Delta Q$

**Table 3.** List of stars with  $\Delta Q > 30$  in HTPM case A, with assumed radial velocities set to zero.

HIP	$\Delta Q$	$H_p$	$\varpi$ [mas]	$\sqrt{\mu_{\alpha*}^2 + \mu_{\delta}^2}$ [mas yr <sup>-1</sup> ]	$v_r$ [km s <sup>-1</sup> ]	$\mu_r$ [mas yr <sup>-1</sup> ]	$\Delta\theta_{\text{persp}}$ [mas]	Remark
87937	8 044.46	9.490	548.31	10 358.94	-110.51	-12 782.22	361.87	Barnard's star
24186	5 053.12	8.932	255.66	8 669.40	245.19	13 223.43	312.01	Kapteyn's star
57939	686.11	6.564	109.99	7 059.03	-98.35	-2 281.95	43.72	Groombridge 1830
104217	618.09	6.147	285.88	5 172.58	-64.07	-3 863.82	54.70	61 Cyg B <sup>a</sup>
54035	572.73	7.506	392.64	4 801.04	-84.69	-7 014.64	92.31	
70890	229.59	10.761	771.64	3 852.57	-22.40	-3 646.21	38.32	Proxima Centauri
74235	76.86	9.200	34.65	3 681.26	310.12	2 266.79	24.80	
439	62.69	8.618	230.42	6 100.36	25.38	1 233.65	20.64	
74234	34.62	9.568	35.14	3 680.96	310.77	2 303.67	21.63	
54211	30.13	8.803	206.27	4 510.10	68.89	2 997.58	36.97	

**Notes.** The threshold 30 was set for a probability of false alarm  $\sim 10^{-5}$ , assuming that  $\Delta Q$  follows the  $\chi^2$  distribution with 5 degrees of freedom. The columns contain the HIPPARCOS identifier, HIPPARCOS magnitude, parallax, and total proper motion (all from the HIPPARCOS Catalogue), the radial velocity from XHIP, and the calculated radial motion and positional offset over  $\Delta t = 23.75$  yr due to perspective acceleration. <sup>(a)</sup> 61 Cyg A (HIP 104214) was not included in the simulations since it is brighter than the nominal *Gaia* bright star limit.

is sensitive to deviations from uniform space motion, whether they are seen in the HIPPARCOS or in the *Gaia* data, or both, this statistic can be used to find candidate systems in all these cases. The further exploration of the candidate systems will, however, require specialised analysis tools.

In a future publication we will explore in more detail how  $\Delta Q$  can be used to identify binary and exoplanetary candidates with orbital periods of decades to centuries. Apart from the possibility to detect substellar companions for the nearest stars, this will contribute to the census of the binary population within a few hundred parsecs from the sun by filling a difficult-to-observe gap between the shorter period spectroscopic and astrometric binaries and the visually resolved long-period systems.

## 5.2. Two versus five parameters

When evaluating the results of our simulations, case B deserves additional attention since it is the more realistic case for the first *Gaia* data release, and the first simulation of this case published so far. The two-parameter solution (*Gaia* 12-B in Table 2) leads to a large position error of several mas. This is caused by assuming the parallax, proper and radial motion to be zero in the solution, whereas in reality they are not. The actual positional uncertainties in this case depend on the true distribution of parallaxes and proper motions for all the stars, including the auxiliary stars, which are not very well known. The numerical values given here are based on the very schematic distribution model for the auxiliary stars described in Sect. 3.2.1, and should therefore be interpreted with caution.

This position error is also relevant for the case B HTPM scenario, where the solution of the auxiliary stars is two parameters only, but where one solves all five parameters for the HIPPARCOS stars while incorporating prior information from the HIPPARCOS Catalogue. The position error of the auxiliary stars causes a poor attitude determination. This in turn leads to increased errors in the case B HTPM results (compare HTPM B and A in Table 2), with a bias in the parallax errors (see Sect. 4 and Fig. 2). For a parallax-unbiased solution it is necessary to estimate all five parameters for all stars included in the solution. Any mixture in the estimation of five and two parameters in the same solution will lead to a bias in the resulting parallaxes. This is not only true for the HTPM scenario described in this paper but also in all *Gaia*-only data releases. Referring to the terminology used in Sect. 6.2

of Lindegren et al. (2012), any star for which not all five astrometric parameters can be solved must be treated as a “secondary source”, meaning that it does not contribute to the attitude determination and instrument calibration. This is necessary in order to avoid biases for the stars where all five parameters are estimated.

## 5.3. Frame rotation of the combined solution

For the final AGIS solution of *Gaia* the reference frame will be established by means of quasars, both by linking to the optical counterparts of radio (VLBI) sources defining the orientation of the International Celestial Reference Frame, and by using the zero proper motion of quasars to determine a non-rotating frame<sup>9</sup>. This can also be done for earlier *Gaia* data releases, at least for the orientation part, while the shorter time span will limit the determination of the spin. It is desirable to rotate the HTPM results into the same reference frame as used for the first *Gaia* data release. This must be done in two steps. First, a provisional HTPM must be computed in the HIPPARCOS frame (as it will be when the HIPPARCOS data are used as prior, see Sect. 2.7), without imposing any other constraints on the frame. This solution will contain (many) non-HIPPARCOS stars with only *Gaia* observations which include a multitude of quasars. Their positions and proper motions are used in a second step to correct the provisional HTPM (and other data in the same solution) for the estimated orientation and spin. Since the HTPM solution is integrated in AGIS, the estimation and correction of the frame can be accomplished using the procedures and tools developed for AGIS (Lindegren et al. 2012, Sect. 6.1).

## 5.4. Other applications of the joint solution method

The joint solution is applicable also to other combinations of astrometric data. Here we give two examples.

Nano-JASMINE (Hatsutori et al. 2009; Yamada et al. 2013) is an ultra-small Japanese satellite, a technology demonstrator for the JASMINE series of near-infrared astrometry missions, scheduled for launch in 2015. It targets bright stars between magnitude 1 and 10, although the exact limits are not yet determined. Based on current performance estimates the uncertainties

<sup>9</sup> The apparent proper motion of quasars due to the Galactocentric acceleration is expected to have an amplitude of  $\sim 4 \mu\text{as yr}^{-1}$  and is taken into account when determining the spin of the reference frame.

in stellar parameters will be similar to or slightly worse than the uncertainties of the HIPPARCOS data. However, the data will still be very valuable since astrometric catalogues are best at their respective epochs and Nano-JASMINE may be the only astrometric mission at its epoch observing the brightest stars in the sky. The Nano-JASMINE data can be analysed together with HIPPARCOS data analogously to the HTPM project to improve the proper motions of bright stars that may not be observed by *Gaia* (Michalik et al. 2013).

The *Tycho-2* Catalogue (Høg et al. 2000) gives positions for 2.5 million stars, derived from starmapper observations of HIPPARCOS. The positions at the reference epoch J1991.25 have a median internal standard error of 7 mas for stars brighter than  $V_T = 9$  mag and 60 mas for the whole catalogue. Combining the *Tycho-2* positions with *Gaia* data using the joint solution scheme would allow us to derive proper motions for these stars with median uncertainties of 0.3 and 2.5 mas yr<sup>-1</sup>, respectively. This is true even in the conservative scenario (*Gaia* 12-B), since the major uncertainty comes from the *Tycho-2* positions. In this combination the proper motions given in *Tycho-2* should not be used, as they may contain systematic errors of a similar magnitude due to the incorporated old photographic material. However, the derived proper motions from a *Tycho-Gaia* Astrometric Solution (TGAS) could be used to correct the photographic positions in order to take advantage of a much longer temporal baseline.

## 6. Conclusions

We have developed the joint solution method for incorporating priors in the core astrometric solution of *Gaia*. The method can be used in the processing of early *Gaia* data to improve the proper motions of the HIPPARCOS stars, the so-called Hundred Thousand Proper Motions project.

Combining astrometric data from very different epochs requires careful treatment of the non-linear effects of the mapping from spherical to rectilinear coordinates and for high velocity stars due to perspective acceleration. Therefore we have introduced a scaled model of kinematics (SMOK), which allows one to handle these effects in a simple and rigorous manner.

Using simulations we have verified that HTPM, using the joint solution method, gives the expected large improvements in proper motion uncertainties for over 100 000 stars in the HIPPARCOS Catalogue. The predicted proper motion uncertainties range from 14 to 134  $\mu\text{s yr}^{-1}$  depending on the amount of *Gaia* data used and the stellar magnitude, about a factor 30 improvement compared with the HIPPARCOS uncertainties.

We have shown that HTPM also delivers improved parallaxes, which, however, may be strongly biased unless a full five-parameter solution can be obtained from *Gaia*-only data also for all non-HIPPARCOS stars. Whether these parallaxes should be published as part of an HTPM release should be decided based on the amount and quality of *Gaia* data available at the time.

The joint solution is applicable also to a combination of *Tycho-2* positions with early *Gaia* data to derive parallaxes and improved proper motions for the 2.5 million stars. We suggest that this possibility of a *Tycho-Gaia* Astrometric Solution (TGAS) should be considered in the *Gaia* data release plan.

The proposed method to calculate HTPM provides a goodness-of-fit measurement  $\Delta Q$  which is sensitive to deviations from the uniform linear space motion. However, accurate radial velocities are required for nearby fast moving stars in order

to avoid mistaking outliers in  $\Delta Q$  for companion signatures. We recommend to publish  $\Delta Q$  as well as the radial velocities used for the HTPM data reduction. This will allow further investigations of outliers which might indicate binary or exoplanetary candidates, and will permit a correction of the HTPM results if better radial velocities become available.

The full power of HTPM will not be reached with the first *Gaia* data, but only in subsequent releases benefiting from the increased sensitivity of  $\Delta Q$  with improved *Gaia* results. Because of the long temporal baseline and the combination of current with historic astrometry, HTPM will remain relevant throughout the final *Gaia* release for the detection and measurement of binary and exoplanetary candidates.

*Acknowledgements.* We thank F. van Leeuwen for clarification on certain data items in the HIPPARCOS Catalogue and for providing valuable feedback as the referee. We also thank C. Fabricius for many useful comments. This work was partly carried out under ESA Contract No. 4000105564/12/NL/GE. Support from the Swedish National Space Board and the Royal Physiographic Society in Lund is gratefully acknowledged.

## References

- Abramowitz, M., & Stegun, I. 2012, Handbook of Mathematical Functions (New York: Dover Publications)
- Anderson, E., & Francis, C. 2012, *Astron. Lett.*, 38, 331
- Bombrun, A., Lindegren, L., Holl, B., & Jordan, S. 2010, A&A, 516, A77
- Brinker, R. C., & Minnick, R. 1995, *The Surveying Handbook*, 2d. edn. (Dordrecht: Kluwer)
- de Bruijne, J., Siddiqui, H., Lammers, U., et al. 2010, in *Relativity in Fundamental Astronomy: Dynamics, Reference Frames, and Data Analysis*, eds. S. A. Klioner, P. K. Seidelmann, & M. H. Soffel, IAU Symp., 261, 331
- de Bruijne, J. H. J., & Eilers, A.-C. 2012, A&A, 546, A61
- Dehnen, W., & Binney, J. J. 1998, *MNRAS*, 298, 387
- Dravins, D., Lindegren, L., & Madsen, S. 1999, A&A, 348, 1040
- Eichhorn, H., & Rust, A. 1970, *Astron. Nachr.*, 292, 37
- ESA 1997, *The HIPPARCOS and Tycho Catalogues* (Noordwijk: ESA), ESA SP, 1200
- Feissel, M., & Mignard, F. 1998, A&A, 331, L33
- Hatsutori, Y., Suganuma, M., Kobayashi, Y., et al. 2009, *Transactions of Space Technology Japan*, 7, 19
- Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27
- Holl, B., Lindegren, L., & Hobbs, D. 2012, A&A, 543, A15
- Lindegren, L., Lammers, U., Hobbs, D., et al. 2012, A&A, 538, A78
- Michalik, D., Lindegren, L., Hobbs, D., Lammers, U., & Yamada, Y. 2012, in *Astronomical Data Analysis Software and Systems XXI*, ed. P. Ballester, D. Egret, & N. P. F. Lorente, ASP Conf. Ser., 461, 549
- Michalik, D., Lindegren, L., Hobbs, D., Lammers, U., & Yamada, Y. 2013, in *Advancing the Physics of Cosmic Distances*, ed. R. de Grijs, IAU Symp., 289, 414
- Mignard, F. 2009, *The Hundred Thousand Proper Motions Project*, Gaia Data Processing and Analysis Consortium (DPAC) technical note GAIA-C3-TN-OCA-FM-040
- Murray, C. A. 1983, *Vectorial astrometry* (Bristol: Adam Hilger)
- Perryman, M. 2014, *The Exoplanet Handbook* (Cambridge, UK: Cambridge University Press)
- Risquez, D., van Leeuwen, F., & Brown, A. G. A. 2013, A&A, 551, A19
- Schlesinger, F. 1917, *AJ*, 30, 137
- Taff, L. G. 1981, *Computational spherical astronomy* (New York: Wiley-Interscience)
- van Altena, W. F. 2013, *Astrometry for Astrophysics* (Cambridge, UK: Cambridge University Press)
- van de Kamp, P. 1977, *Vistas in Astronomy*, 21, 289
- van Leeuwen, F. 2007a, HIPPARCOS, the New Reduction of the Raw Data (Springer), *Astrophys. Space Sci. Libr.*, 350
- van Leeuwen, F. 2007b, A&A, 474, 653
- Wilson, E. B., & Hilferty, M. M. 1931, *Proc. of the National Academy of Science*, 17, 684
- Yamada, Y., Fujita, S., Gouda, N., et al. 2013, in *Advancing the Physics of Cosmic Distances*, ed. R. de Grijs, IAU Symp., 289, 429

## Appendix A: Scaled modelling of kinematics (SMOK)

A formalism called SMOK is introduced in this paper to facilitate a rigorous manipulation of small (differential) quantities in the celestial coordinates. It is reminiscent of the “standard” or “tangential” coordinates in classical small-field astrometry (e.g., Murray 1983; van Altena 2013), using a gnomonic projection onto a tangent plane of the (unit) celestial sphere, but extends to three dimensions by adding the radial coordinate perpendicular to the tangent plane. This simplifies the modelling of perspective effects.

Figure A.1 illustrates the concept. In the vicinity of the star let  $c$  be a comparison point fixed with respect to the solar system barycentre (SSB). As shown in the diagrams:

1. The barycentric motion of the star is scaled by the inverse distance to  $c$ , effectively placing the star on or very close to the unit sphere.
2. Rectangular coordinates are expressed in the barycentric  $[p_c q_c r_c]$  system with  $r_c$  pointing towards  $c$ , and  $p_c, q_c$  in the directions of increasing right ascension and declination.

The first point eliminates the main uncertainty in the kinematic modelling of the star due to its poorly known distance. The second point allows us to express the scaled kinematic model in SMOK coordinates  $a, d, r$  that are locally aligned with  $\alpha, \delta$ , and the barycentric vector.

Up to the scale factor  $|c|^{-1}$  discussed below, the SMOK coordinate system is completely defined by the adopted comparison point  $(\alpha_c, \delta_c)$  using the orthogonal unit vectors

$$p_c = \begin{bmatrix} -\sin \alpha_c \\ \cos \alpha_c \\ 0 \end{bmatrix}, \quad q_c = \begin{bmatrix} -\sin \delta_c \cos \alpha_c \\ -\sin \delta_c \sin \alpha_c \\ \cos \delta_c \end{bmatrix}, \quad r_c = \begin{bmatrix} \cos \delta_c \cos \alpha_c \\ \cos \delta_c \sin \alpha_c \\ \sin \delta_c \end{bmatrix}. \quad (\text{A.1})$$

$[p_c q_c r_c]$  is the “normal triad” at the comparison point with respect to the celestial coordinate system (Murray 1983)<sup>10</sup>. We are free to choose  $(\alpha_c, \delta_c)$  as it will best serve our purpose, but once chosen (for a particular application) it is fixed: it has no proper motion, no parallax, and no associated uncertainty. Typically  $(\alpha_c, \delta_c)$  is chosen very close to the mean position of the star.

The motion of the star in the Barycentric Celestial Reference System (BCRS) is represented by the function  $b(t)$ , where  $b$  is the vector from SSB to the star as it would be observed from the SSB at time  $t$ . The scaled kinematic model  $s(t) = b(t)|c|^{-1}$  is given in SMOK coordinates as

$$a(t) = p'_c s(t), \quad d(t) = q'_c s(t), \quad r(t) = r'_c s(t), \quad (\text{A.2})$$

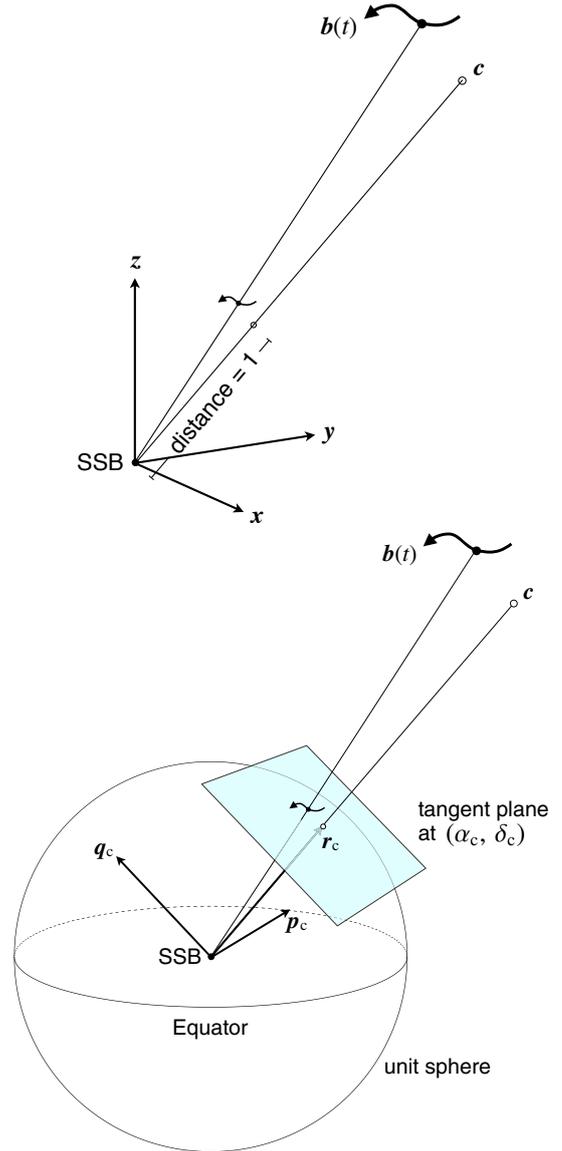
and can in turn be reconstructed from the SMOK coordinates as

$$s(t) = p_c a(t) + q_c d(t) + r_c r(t). \quad (\text{A.3})$$

$a, d, r$  are dimensionless and the first two are typically small quantities ( $\lesssim 10^{-4}$ ), while  $r$  is very close to unity.

The whole point of the scaled kinematic modelling is that  $s(t)$  can be described very accurately by astrometric observations, even though  $b(t)$  may be poorly known due to a high uncertainty in distance. This is possible simply by choosing the

<sup>10</sup>  $p_c$  and  $q_c$  point to the local “East” and “North”, respectively, provided that  $|\delta_c| < 90^\circ$ . However, the coordinate triad in Eq. (A.1) is well-defined even *exactly* at the poles, where  $\alpha_c$  remains significant for defining  $p_c$  and  $q_c$ .



**Fig. A.1.** Two steps in the definition of SMOK coordinates. In the *top* diagram the motion of an object in the vicinity of the fixed point  $c$  is modelled by the function  $b(t)$  expressed in the barycentric  $[x y z]$  system. A scaled version of the model is constructed such that the scaled  $c$  is at unit distance from the solar system barycentre (SSB). In the *bottom* diagram new coordinate axes  $[p_c q_c r_c]$  are chosen in the directions of increasing right ascension, declination, and distance, respectively, at the comparison point  $(\alpha_c, \delta_c)$  being the projection of  $c$  on the unit sphere.

scaling such that  $|s(t)| = 1$  at some suitable time. This works even if the distance is completely unknown, or if it is effectively infinite (as for a quasar).

The scale factor is  $|c|^{-1} = \varpi_c/A$ , where  $\varpi_c$  is the parallax of  $c$  and  $A$  the astronomical unit. The measured parallax can be regarded as an estimate of  $\varpi_c$ .

In the following we describe some typical applications of SMOK coordinates.

### A.1. Uniform space motion

The simplest kinematic model is to assume that the star moves uniformly with respect to the SSB, that is

$$b(t) = b_{ep} + (t - t_{ep})v, \quad (\text{A.4})$$

where  $\mathbf{b}_{\text{ep}}$  is the barycentric position at the reference epoch  $t_{\text{ep}}$ , and  $\mathbf{v}$  is the (constant) space velocity. The scaled kinematic model expressed in the BCRS is

$$\mathbf{s}(t) = \mathbf{s}_{\text{ep}} + (t - t_{\text{ep}})\dot{\mathbf{s}}, \quad (\text{A.5})$$

where

$$\mathbf{s}_{\text{ep}} = \mathbf{p}_c \mathbf{a}(t_{\text{ep}}) + \mathbf{q}_c \mathbf{d}(t_{\text{ep}}) + \mathbf{r}_c r(t_{\text{ep}}) \quad (\text{A.6})$$

and

$$\dot{\mathbf{s}} = \mathbf{p}_c \dot{\mathbf{a}} + \mathbf{q}_c \dot{\mathbf{d}} + \mathbf{r}_c \dot{r} \quad (\text{A.7})$$

are constant vectors. The uniform motion can also be written in SMOK coordinates as

$$\left. \begin{aligned} a(t) &= a(t_{\text{ep}}) + (t - t_{\text{ep}})\dot{a}, \\ d(t) &= d(t_{\text{ep}}) + (t - t_{\text{ep}})\dot{d}, \\ r(t) &= r(t_{\text{ep}}) + (t - t_{\text{ep}})\dot{r}. \end{aligned} \right\} \quad (\text{A.8})$$

The six constants  $a(t_{\text{ep}})$ ,  $d(t_{\text{ep}})$ ,  $r(t_{\text{ep}})$ ,  $\dot{a}$ ,  $\dot{d}$ ,  $\dot{r}$  are the kinematic parameters of the scaled model; however, to get the actual kinematics of the star we also need to know  $\varpi_c$ .

## A.2. Relation to the usual astrometric parameters

Choosing  $(\alpha_c, \delta_c)$  to be the barycentric celestial coordinates of the star at  $t_{\text{ep}}$ , and  $\varpi_c$  equal to the parallax at the same epoch, we find

$$\left. \begin{aligned} a(t_{\text{ep}}) &= 0, & d(t_{\text{ep}}) &= 0, & r(t_{\text{ep}}) &= 1, \\ \dot{a} &= \mu_{\alpha^*}, & \dot{d} &= \mu_{\delta}, & \dot{r} &= \mu_r, \end{aligned} \right\} \quad (\text{A.9})$$

where  $\mu_{\alpha^*}$ ,  $\mu_{\delta}$  are the tangential components of the barycentric proper motion at the reference epoch  $t_{\text{ep}}$ , and  $\mu_r$  is the ‘‘radial proper motion’’ allowing one to take the perspective effects into account.  $\mu_r$  is usually calculated from the measured radial velocity and parallax according to Eq. (1).

## A.3. Differential operations

Uniform space motion does not map into barycentric coordinates  $\alpha(t)$ ,  $\delta(t)$  that are linear functions of time. The non-linearity derives both from the curvilinear nature of spherical coordinates and from perspective foreshortening depending on the changing distance to the object. Both effects are well known and have been dealt with rigorously by several authors (e.g., [Eichhorn & Rust 1970](#); [Taff 1981](#)). The resulting expressions are non-trivial and complicate the comparison of astrometric catalogues of different epochs. For example, approximations such as

$$\mu_{\alpha^*} = \frac{\alpha(t_2) - \alpha(t_1)}{t_2 - t_1} \cos \delta, \quad \mu_{\delta} = \frac{\delta(t_2) - \delta(t_1)}{t_2 - t_1} \quad (\text{A.10})$$

cannot be used when the highest accuracy is required. By contrast, the linearity of Eq. (A.8) makes it possible to write

$$\dot{a} = \frac{a(t_2) - a(t_1)}{t_2 - t_1}, \quad \dot{d} = \frac{d(t_2) - d(t_1)}{t_2 - t_1} \quad (\text{A.11})$$

to full accuracy, provided that the same comparison point is used for both epochs. (Strictly speaking, the same scale factor must also be used, so that in general  $r(t_2) - r(t_1) = (t_2 - t_1)\dot{r} \neq 0$ .) If the position at the reference epoch coincides with the comparison point used, the resulting  $\dot{a}$ ,  $\dot{d}$  are the looked-for proper motion components according to Eq. (A.9); otherwise a change of comparison point is needed (see below).

## A.4. Changing the comparison point

Let  $(\alpha_1, \delta_1)$  and  $(\alpha_2, \delta_2)$  be different comparison points with associated triads  $[\mathbf{p}_1 \mathbf{q}_1 \mathbf{r}_1]$  and  $[\mathbf{p}_2 \mathbf{q}_2 \mathbf{r}_2]$ . If  $a_1(t)$ ,  $d_1(t)$ ,  $r_1(t)$  and  $a_2(t)$ ,  $d_2(t)$ ,  $r_2(t)$  describe the same scaled kinematics we have by Eq. (A.3)

$$\mathbf{s}(t) = \mathbf{p}_1 a_1(t) + \mathbf{q}_1 d_1(t) + \mathbf{r}_1 r_1(t) = \mathbf{p}_2 a_2(t) + \mathbf{q}_2 d_2(t) + \mathbf{r}_2 r_2(t). \quad (\text{A.12})$$

Thus, given  $a_1(t)$ ,  $d_1(t)$ ,  $r_1(t)$  one can compute  $\mathbf{s}(t)$  from the first equality in Eq. (A.12), whereupon the modified functions are recovered as

$$a_2(t) = \mathbf{p}'_2 \mathbf{s}(t), \quad d_2(t) = \mathbf{q}'_2 \mathbf{s}(t), \quad r_2(t) = \mathbf{r}'_2 \mathbf{s}(t). \quad (\text{A.13})$$

This procedure can be applied to  $\mathbf{s}(t)$  for any particular  $t$  as well as to linear operations on  $\mathbf{s}$  such as differences and time derivatives.

## A.5. Epoch propagation

An important application of the above formulae is for propagating the six astrometric parameters  $(\alpha_1, \delta_1, \varpi_1, \mu_{\alpha^*1}, \mu_{\delta1}, \mu_{r1})$ , referring to epoch  $t_1$ , to a different epoch  $t_2$ . This can be done in the following steps:

1. Use  $(\alpha_1, \delta_1)$  as the comparison point and compute  $[\mathbf{p}_1 \mathbf{q}_1 \mathbf{r}_1]$  by Eq. (A.1). At time  $t_1$  the SMOK parameters relative to the first comparison point are  $a_1(t_1) = d_1(t_1) = 0$ ,  $r_1(t_1) = 1$ ,  $\dot{a}_1 = \mu_{\alpha^*1}$ ,  $\dot{d}_1 = \mu_{\delta1}$ ,  $\dot{r}_1 = \mu_{r1}$ .
2. Calculate  $\mathbf{s}(t_1)$  and  $\dot{\mathbf{s}}$  using Eqs. (A.6)–(A.7).
3. Calculate  $\mathbf{s}(t_2)$  by means of Eq. (A.5). Let  $s_2 = |\mathbf{s}(t_2)|$  be its length (close to unity).
4. Calculate  $\mathbf{r}_2 = \mathbf{s}(t_2)/s_2$  and hence the second comparison point  $(\alpha_2, \delta_2)$  and triad  $[\mathbf{p}_2 \mathbf{q}_2 \mathbf{r}_2]$ .
5. Use Eq. (A.13) to calculate the SMOK parameters at  $t_2$  referring to the second comparison point. For the position one trivially gets  $a_2(t_2) = d_2(t_2) = 0$  and  $r_2(t_2) = s_2$ . For the proper motion parameters one finds  $\dot{a}_2 = \mathbf{p}'_2 \dot{\mathbf{s}}$ ,  $\dot{d}_2 = \mathbf{q}'_2 \dot{\mathbf{s}}$ , and  $\dot{r}_2 = \mathbf{r}'_2 \dot{\mathbf{s}}$ .
6. The astrometric parameters at epoch  $t_2$  are  $\alpha_2, \delta_2, \varpi_2 = \varpi_1/s_2, \mu_{\alpha^*2} = \dot{a}_2/s_2, \mu_{\delta2} = \dot{d}_2/s_2, \mu_{r2} = \dot{r}_2/s_2$ .

This procedure is equivalent to the one described in Sect. 1.5.5, Vol. 1 of *The HIPPARCOS and Tycho Catalogues* ([ESA 1997](#)).

## Appendix B: The HIPPARCOS Catalogue

This Appendix describes the calculation of relevant quantities from the new reduction of the HIPPARCOS Catalogue by [van Leeuwen \(2007b\)](#). Data files were retrieved from the Strasbourg astronomical Data Center (CDS) in November 2013 (catalogue I/311). These files differ slightly from the ones given on the DVD published along with the book ([van Leeuwen 2007a](#)), both in content and format, as some errors have been corrected. The data needed for every accepted catalogue entry are:

- the five astrometric parameters  $(\alpha, \delta, \varpi, \mu_{\alpha^*}, \mu_{\delta})$ ;
- the  $5 \times 5$  normal matrix  $N$  from the least-squares solution of the astrometric parameters (for a 5-parameter solution this equals the inverse of the covariance matrix  $C$ );
- the chi-square goodness-of-fit quantity  $Q$  for the 5-parameter solution of the HIPPARCOS data;
- the degrees of freedom  $\nu$  associated with  $Q$ .

The astrometric parameters at the HIPPARCOS reference epoch J1991.25 are directly taken from the fields labelled RARad, DErad, Plx, pmRA, and pmDE in the main catalogue file `hip2.dat`. Units are [rad] for  $\alpha$  and  $\delta$ , [mas] for  $\varpi$ , and [mas yr<sup>-1</sup>] for  $\mu_{\alpha^*}$  and  $\mu_\delta$ . It is convenient to express also positional differences (such as SMOK coordinates  $a$  and  $d$ ) and positional uncertainties in [mas]. The elements of  $N$  thus have units [mas<sup>-2</sup> yr <sup>$p$</sup> ], where  $p = 0, 1$ , or  $2$ , depending on the position of the element in the matrix.

The calculation of  $N$ ,  $Q$ , and  $\nu$  is described hereafter in some detail as the specification of  $C$  deviates in some details from the published documentation. Clarification on certain issues was kindly provided by van Leeuwen (priv. comm.).

The number of degrees of freedom is

$$\nu = N_{\text{tr}} - n, \quad (\text{B.1})$$

where  $N_{\text{tr}}$  is the number of field transits used (label `Ntr` in `hip2.dat`) and  $n$  is the number of parameters in the solution (see below; most stars have  $n = 5$ ). The goodness-of-fit given in field F2 is the ‘‘gaussianized’’ chi-square (Wilson & Hilferty 1931)

$$F_2 = \left(\frac{9\nu}{2}\right)^{1/2} \left[ \left(\frac{Q}{\nu}\right)^{1/3} + \frac{2}{9\nu} - 1 \right] \quad (\text{B.2})$$

computed from  $Q$ , the sum of the squared normalized residuals, and  $\nu$ . For ‘‘good’’ solutions  $Q$  is expected to follow the chi-square distribution with  $\nu$  degrees of freedom ( $Q \sim \chi^2(\nu)$ ), in which case  $F_2$  approximately follows the standard normal distribution,  $F_2 \sim N(0, 1)$ . Thus,  $F_2 > 3$  means that  $Q$  is ‘‘too large’’ for the given  $\nu$  at the same level of significance as the  $+3\sigma$  criterion for a Gaussian variable (probability  $\lesssim 0.0044$ )<sup>11</sup>. Given  $F_2$  from field F2, and  $\nu$  from Eq. (B.1), it is therefore possible to reconstruct the chi-square statistic of the  $n$ -parameter solution as

$$Q = \nu \left[ \left(\frac{2}{9\nu}\right)^{1/2} F_2 + 1 - \frac{2}{9\nu} \right]^3. \quad (\text{B.3})$$

We also introduce the square-root of the reduced chi-square,

$$u = \sqrt{Q/\nu}, \quad (\text{B.4})$$

which is expected to be around 1.0 for a ‘‘good’’ solution (see further discussion below).  $u$  is sometimes referred to as the standard error of unit weight (Brinker & Minnick 1995).

The catalogue gives the covariance matrix in the form of an upper-diagonal ‘‘weight matrix’’  $U$  such that, formally,  $C = (U'U)^{-1}$ . This inverse exists for all stars where a solution is given. (For the joint solution we actually need the normal matrix  $N = U'U$ , see below.) For solutions with  $n = 5$  astrometric parameters there are  $n(n + 1)/2 = 15$  non-zero elements in  $U$ . For some stars the solution has more than five parameters, and the main catalogue then only gives the first 15 non-zero elements, while remaining elements are given in separate tables.

Let  $U_1, U_2, \dots, U_{15}$  be the 15 values taken from the fields labelled `UW` in `hip2.dat`. The matrix  $U$  is computed as

$$U = \begin{bmatrix} f_1 U_1 & U_2 & U_4 & U_7 & U_{11} \\ 0 & f_2 U_3 & U_5 & U_8 & U_{12} \\ 0 & 0 & f_3 U_6 & U_9 & U_{13} \\ 0 & 0 & 0 & f_4 U_{10} & U_{14} \\ 0 & 0 & 0 & 0 & f_5 U_{15} \end{bmatrix}. \quad (\text{B.5})$$

Here  $f_i, i = 1 \dots n$ , are scaling factors which for the CDS data must be calculated as

$$\begin{aligned} f_1 &= u/\sigma_{\alpha^*}, & f_2 &= u/\sigma_\delta, & f_3 &= u/\sigma_\varpi, \\ f_4 &= u/\sigma_{\mu_{\alpha^*}}, & f_5 &= u/\sigma_{\mu_\delta}, \end{aligned} \quad (\text{B.6})$$

where  $u$  is given by Eq. (B.4) and  $\sigma_i$  are the standard errors given in fields `e_RARad` through `e_pmDE` of `hip2.dat`. Equation (B.6) applies to data taken from the CDS version of the catalogue (I/311). For catalogue data on the DVD accompanying the book (van Leeuwen 2007a), scaling factors  $f_i = 1$  apply, although those data are superseded by the CDS version.

The  $5 \times 5$  matrix  $N = U'U$  computed using the first five rows and columns in  $U$ , as given in Eq. (B.5), contains the relevant elements of the normal matrix for any solution with  $n \geq 5$ . Thus, for solutions with  $n > 5$  there is no need, for the catalogue combination, to retrieve the additional elements of  $U$  from `hip7p.dat`, etc. The situation is different when the covariance matrix is needed: it is then necessary to compute the full  $n \times n$  normal matrix  $N$  before  $C = N^{-1}$  can be computed.

The normal matrix  $N$  computed as described above incorporates the formal uncertainties of the observations; as described in van Leeuwen (2007a) these are ultimately derived from the photon statistics of the raw data after careful analysis of the residuals as function of magnitude, etc. If the adopted models are correct we expect the  $F_2$  statistic to be normally distributed with zero mean and unit standard deviation, and the standard error of unit weight,  $u$ , to be on the average equal to 1. In reality we find (for solutions with  $n = 5$ ) that their distributions are skewed towards higher values, especially for the bright stars where photon noise is small and remaining calibration errors are therefore relatively more important. To account for such additional errors the published standard errors  $\sigma_{\alpha^*}$ , etc., in `hip2.dat` include, on a star-by-star basis, a correction factor equal to the unit weight error  $u$  obtained in its solution. This is equivalent to scaling the formal standard errors of the data used in the solution by the same factor. In order to make the computed normal matrix, covariance matrix, and goodness-of-fit statistics consistent with the published standard errors it is then necessary to apply the corresponding corrections, viz.:

$$N_{\text{corr}} = N u^{-2}, \quad C_{\text{corr}} = C u^2, \quad Q_{\text{corr}} = \nu, \quad u_{\text{corr}} = 1. \quad (\text{B.7})$$

For the catalogue combination we use  $N_{\text{corr}}$  and  $Q_{\text{corr}}$  whenever  $u > 1$ , but  $N$  and  $Q$  if  $u \leq 1$ .

<sup>11</sup> This transformation was also used to generate the F2 statistic given in field H30 of the HIPPARCOS and *Tycho* Catalogues (ESA 1997).