Estimating distances from parallaxes. VI:
A method for inferring distances and transverse velocities from parallaxes and proper motions demonstrated on Gaia Data Release 3

C.A.L. BAILER-JONES[1]

[1] *Max Planck Institute for Astronomy, Heidelberg, Germany*

ABSTRACT

The accuracy of stellar distances inferred purely from parallaxes degrades rapidly with distance. Proper motion measurements, when combined with some idea of typical velocities, provide independent information on stellar distances. Here I build a direction- and distance-dependent model of the distribution of stellar velocities in the Galaxy, then use this together with parallaxes and proper motions to infer kinegeometric distances and transverse velocities for stars in Gaia DR3. Using noisy simulations I assess the performance of the method and compare its accuracy to purely parallax-based (geometric) distances. Over the whole Gaia catalogue, kinegeometric distances are on average 1.25 times more accurate than geometric ones. This average masks a large variation in the relative performance, however. Kinegeometric distances are considerably better than geometric ones beyond several kpc, for example. On average, kinegeometric distances can be measured to an accuracy of 19% and velocities ($\sqrt{v_{\alpha*}^2 + v_\delta^2}$) to $16\,\mathrm{km\,s^{-1}}$ (median absolute deviations). In Gaia DR3, kinegeometric distances are smaller than geometric ones on average for distant stars, but the pattern is more complex in the bulge and disk. With the much more accurate proper motions expected in Gaia DR5, a further improvement in the distance accuracy by a factor of (only) 1.35 on average is predicted (with kinegeometric distances still 1.25 times more accurate than geometric ones). The improvement attained from proper motions is limited by the width of the velocity prior, in a way that the improvement from better parallaxes is not limited by the width of the distance prior.

1. INTRODUCTION

Precise parallaxes may be simply inverted to get precise distance estimates of individual stars. This is not possible for low precision parallaxes, however, because the inversion induces a disproportionately large random error and bias (Bailer-Jones 2015; Luri et al. 2018). Of the 1.47 billion stars with parallaxes in Gaia Data Release 3 (GDR3) (Gaia Collaboration et al. 2023), 87% have a parallax SNRs less than five, and 78% less than three. Even in the final Gaia data release the parallax SNRs will only improve by a factor of two for most stars, leaving 75% still with parallax SNRs less than five (60% less than three).

A more considered approach than parallax inversion is therefore required for distance estimation. This has been framed as a probabilistic inference (Bayesian) problem in the present series of papers. In the first paper (Bailer-Jones 2015) I looked at the general problem of inferring distances from parallaxes, and introduced the exponentially-decreasing space density (EDSD) prior. The second paper (Astraatmadja & Bailer-Jones 2016a) introduced a more sophisticated distance prior based on a model of the Milky Way. Both this and simpler isotropic priors were applied in the third paper (Astraatmadja & Bailer-Jones 2016b) to estimate distances for all two million stars in GDR1 that had parallaxes. Paper four (Bailer-Jones et al. 2018) built a direction-dependent EDSD prior for the whole Galaxy using a mock Gaia catalogue and used this to infer geometric distances for all 1.33 billion stars in GDR2 with parallaxes. Bailer-Jones et al. (2021), the fifth in the series, did this again for the higher precision parallaxes in GeDR3 using a more flexible direction-dependent distance prior (the generalized gamma distribution). This paper also extended the method to take advantage of the Gaia colour and apparent magnitude of each star which,

together with a model for the colour–absolute magnitude diagram (which also varied over the sky), gives information on distance (as already explored in Astraatmadja & Bailer-Jones 2016a). These distances, which we called *photogeometric distances*, were published for 1.35 billion stars, along with geometric distances for 1.47 billion stars.

Several other authors have published distance estimation methods or catalogues using the Gaia parallaxes. Some focus on specific objects, such as Wolf-Rayet stars (Rate & Crowther 2020) or Mira variables (Sanders 2023) or stellar clusters (Cantat-Gaudin et al. 2020; Olivares et al. 2020). Others have performed large scale distance estimation of broad types of individual stars for which more general priors must be used. Both Fouesneau et al. (2022) and Anders et al. (2022) combined Gaia parallaxes with optical and infrared photometry from multiple surveys to infer intrinsic stellar properties, line-of-sight extinctions, and distances for over a hundred million stars. Andrae et al. (2023) used the Gaia BP/RP spectra and parallaxes to also estimate stellar parameters, extinctions, and distances, and published these as part of GDR3 for half a billion stars. Using the publicly-released subset of the GDR3 BP/RP spectra, Zhang et al. (2023) also estimated stellar parameters (albeit parallax rather than distance, so the inversion problem remains). Although conceptually similar to the approach used to produce photogeometric distances in Bailer-Jones et al. (2021), these large-scale distance estimation projects differ in their methods, and by using the spectral energy distribution make stronger assumptions about the stars themselves.

Stellar proper motions can provide further distance information. A star moving with an angular velocity μ at distance r has a transverse velocity of

$$\frac{v}{\mathrm{km\,s^{-1}}} = k\,\frac{\mu}{\mathrm{mas\,yr^{-1}}}\,\frac{r}{\mathrm{kpc}} \quad \text{with} \quad k = 4.740471 \ . \quad (1)$$

Assuming stars have a limited range of plausible velocities, a measurement of the proper motion puts a constraint on the plausible distance. A velocity prior quantifies what we mean by "plausible" and inference delivers a posterior over distance. Alone this would provide a rather poor distance estimate, but when combined with a parallax measurement we achieve a *kinegeometric distance* estimate that is potentially more accurate than a purely geometric one. This is of particular interest for distant stars in Gaia, as these tend to have very low SNR parallaxes but still relatively accurate proper motions. I explore this approach in this paper.

The use of proper motions to aid distance estimation is not new (see Binney & Merrifield 1998, section 2.2 for an overview). A simple case is to adopt a model of circu-

lar rotation in the Galaxy (a rather narrow prior), and to estimate the distance directly from the proper motion (e.g. Reid 2022). Proper motions have long been used to estimate the mean distance to a group of stars, often under the name of statistical parallaxes, by measuring the radial velocities as well as the proper motions (e.g. Strömberg 1936, Hemenway 1975). Schönrich et al. (2019) adopted a related approach to reduce the bias in parallax-based distance estimates, and used this to estimate distances to 7 million stars with RVs in GDR2. Zari et al. (2021) used a similar approach to the one in the present paper to estimate distances to OBA stars in the Galactic disk from GeDR3. As the target population was narrowly defined, the distance and velocity priors could be tailored to the expected distribution for young stars out to 3–4 kpc.

In the present paper I use a direction and distant-dependent velocity prior derived from a Galaxy model to estimate kinegeometric distances to a random subset of stars from the entire GDR3 catalogue (Gaia Collaboration 2022). I compare the results to geometric distances, and assess the performance of both types of estimate using a mock catalogue. The kinegeometric method also provides estimates of the two transverse velocity components. Radial velocities are not used: When I refer to "velocity" in this paper it means the total transverse velocity, not the 3D velocity. In contrast to the previous three papers in this series, I do not publish a catalogue because, as we shall see, the improvements from using proper motions are substantial only in a limited part of parameter space.

2. METHOD

The prior is constructed for each HEALpixel (Górski et al. 2005) independently. As in Bailer-Jones et al. (2021), I use the equatorial, nested scheme at level 5, giving 12 288 HEALpixels with an area of 3.36 square degrees each. All sky plots in this paper are a Mollweide equal-area projection in Galactic coordinates, with north up, east increasing to the left, and $(l, b) = (0°, 0°)$ in the centre. They were plotted using *healpy* (Zonca et al. 2019) and show grid lines at 45° in latitude and longitude. All other plots, as well as the processing and analysis, were done in R (r-project.org).

2.1. *Distance and transverse velocity posterior*

Bailer-Jones et al. (2021) estimated geometric distances using a prior that varied only with distance (r) for each HEALpixel (p). The product of this prior and the likelihood is the unnormalized posterior probability

distribution function (PDF)

$$P_{\mathrm{g}}^{*}(r \mid \varpi, \sigma_{\varpi}, p) \ = \ P(\varpi \mid r, \sigma_{\varpi}) \, P(r \mid p) \ . \quad (2)$$

The likelihood is a Gaussian in parallax (ϖ) with uncertainty σ_{ϖ}, as defined in section 2.2 of Bailer-Jones et al. (2021). For comparison purposes I below report geometric distances inferred from this posterior. These use the same input data, parallax zeropoint, and prior as in Bailer-Jones et al. (2021), so are statistically identical.

Here I introduce the *kinegeometric distance*, which uses the proper motions ($\mu_{\alpha*}, \mu_{\delta}$) in addition to the parallax. Proper motions are connected to the distance via equation 1, so to exploit them we must introduce both a prior on the velocities ($v_{\alpha*}, v_{\delta}$) and a likelihood on the proper motions. Combining all this we get the 3D kinegeometric posterior PDF over distance and velocity

$$P_{\mathrm{kg}}^{*}(r, v_{\alpha*}, v_{\delta} \mid \varpi, \mu_{\alpha*}, \mu_{\delta}, \Sigma, p) \ = \\ \underbrace{P(\varpi, \mu_{\alpha*}, \mu_{\delta} \mid r, v_{\alpha*}, v_{\delta}, \Sigma)}_{\text{likelihood}} \ \underbrace{P(r, v_{\alpha*}, v_{\delta} \mid p)}_{\text{prior}} \ , \quad (3)$$

the * symbol indicating that the PDF is unnormalized. The likelihood is a 3D Gaussian with mean

$$\left(\frac{1}{r}, \frac{v_{\alpha*}}{kr}, \frac{v_{\delta}}{kr} \right) \qquad (4)$$

and covariance Σ, which is the 3×3 variance–covariance matrix of the parallax and proper motions published in GDR3 for each star (Lindegren et al. 2021a; Gaia Collaboration et al. 2023). The form of the likelihood shows that the more precise the measured proper motion, the more precise the estimated value of v/r. As with the geometric distances, I use the parallax zeropoint of Lindegren et al. (2021b). I make no modifications to the proper motions. The prior will be introduced below. From the 3D posterior we can estimate the distance and the transverse velocity components by marginalization. As I numerically sample the posterior (section 2.3), this marginalization is trivial.

The main goal of this work is to assess the improvement in distance estimates by using the proper motions in addition to the parallaxes, but I will also investigate the velocities estimated by equation 3. Transverse velocities can alternatively be obtained by multiplying the proper motions by the geometric distance. These I refer to as *geometric* velocities, and will compute them for comparison purposes. I do not consider estimating velocities by dividing the proper motion by the parallax: this is very biased, and fails completely when the parallax is not positive, which is the case for 24% of sources with parallaxes in GDR3.
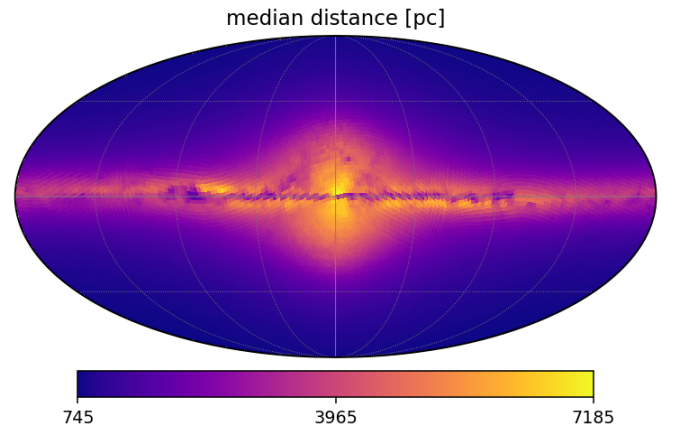

Figure 1. Median of the distance prior in each HEALpixel, computed from the mock catalogue of Rybizki et al. (2020).

2.2. *Velocity prior*

The distance–velocity prior in equation 3 can be decomposed without loss of generality as

$$P(r, v_{\alpha*}, v_{\delta} \mid p) \ = \ P(v_{\alpha*}, v_{\delta} \mid r, p) \, P(r \mid p) \ . \quad (5)$$

For the second term – the distance prior – I adopt the same one used for the geometric distances. This is described in (Bailer-Jones et al. 2021, section 2.3). It was built from the mock Gaia catalogue of Rybizki et al. (2020) and is summarized in Figure 1. The first term in equation 5, the velocity prior, I compute here from the same mock catalogue, which is based on the Galaxy model of Robin et al. (2003) plus some later modifications. This model comprises several Galactic populations – bulge, thin and thick disks of a range of ages, stellar halo, dark matter halo, interstellar medium – with velocities computed self-consistently with the mass distribution via age–velocity relations. The disk model includes a warp and flare and the outer bulge is strongly oblate (a bar). I exclude all star clusters.

Stellar velocity distributions vary considerably with direction and distance, so a dependence on both is retained. The prior should be a continuous – and ideally smooth –function of distance and direction. Here I compute it independently for each HEALpixel (direction) so it is not smooth across HEALpixel boundaries, but it is smooth in distance. For a given HEALpixel, I investigated how the velocity distribution varied over distance shells, a shell being defined as the volume between two spheres of different radii centered on the Sun. With sufficient stars, the 2D distribution in each shell could be approximated well with a 2D Gaussian.

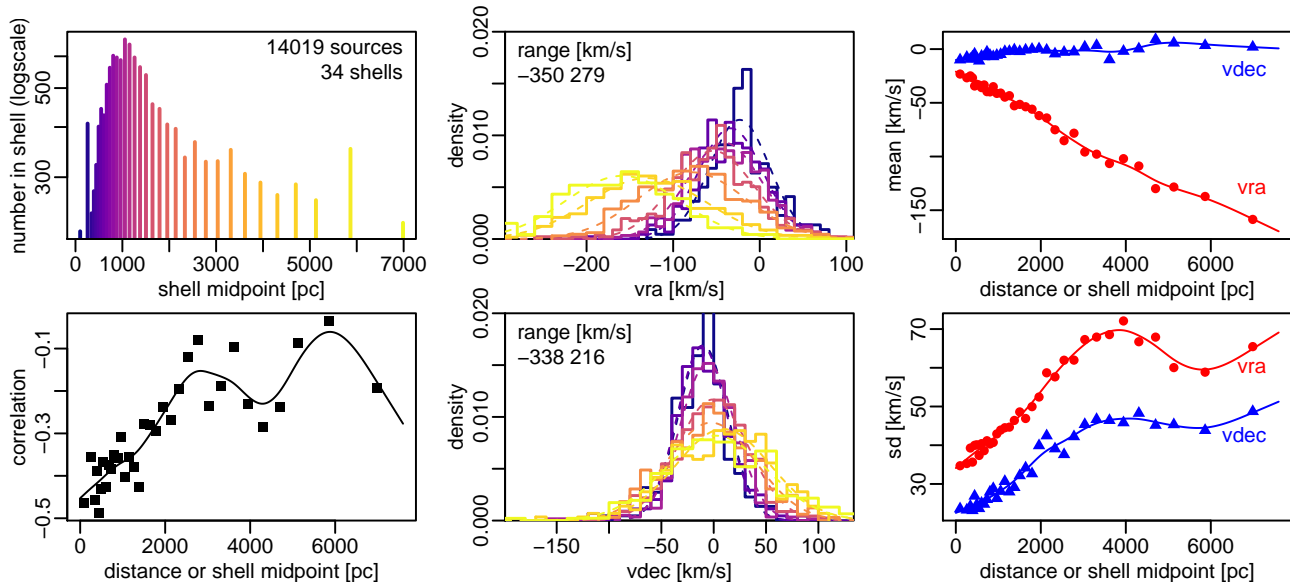The task of building a prior then falls to determining how the mean and variance–covariance of this2D Gaus-

Figure 2. Velocity prior for HEALpixel 6200 at $(l, b) = (285.7°, 34.8°)$. The top left panel shows the positions of the mid points of the shells and the number of sources in them (on a log scale). The two central panels show the distribution of the velocities (top $v_{\alpha*}$, bottom v_δ) for some selected shells, with the colours denoting the shell as in the top left panel. The dashed lines show the fitted Gaussians. The two right panels show the values of the fitted means (upper) and fitted standard deviations (lower) for all shells, as red circles for $v_{\alpha*}$ and as blue triangles for v_δ. The lines show the spline fits, which extend to the outer radius of the last shell, here 7597 pc. The bottom left panel shows the fitted correlations of the Gaussians in each shell along with its spline fit, but this is not used (the correlation is set to zero in the model).

sian should vary with distance. Figure 2 supports the following explanation. I first construct a series of contiguous shells moving outwards from the Sun, defined such that the volume of each shell is a fixed multiple λ of the volume of the previous shell. When the outer radius of the first shell is fixed (its inner radius is zero), this recurrence relation defines all shells. If stars were uniformly distributed in volume, then $\lambda = 1$ would produce shells with the same number of stars.

After experimenting to achieve a trade-off between having sufficiently small shells to capture the variation of the velocity distribution, yet sufficiently large shells to allow an accurate Gaussian fit, I selected $\lambda = 1.3$ and an initial outer shell radius of 200 pc. I generate a set of 45 shells such that the outer radius of the last shell is 15.3 kpc. Stars beyond this do not contribute to the prior. Gaia astrometry on more distant stars is poor, so larger distances are not worth accommodating. Some HEALpixels at high Galactic latitudes still have too few stars, so I merge successive shells (starting from the Sun) to attain at least 200 stars per shell. At higher latitudes this leads to the outermost shell having an outer radius less than 15.3 kpc (as there are fewer than 200 stars beyond). The smallest outermost radius is 3.2 kpc with a median of 9.0 kpc. Among the 12 228 HEALpixels, the number of shells varies from 11 to 45 with a median value of 36. The number of stars per HEALpixel varies

from 3200 to 3.3 million with a median of 18 thousand. The top-left panel of Figure 2 shows an example of the shell structure for one HEALpixel.

Once the shells have been fixed, I compute robust estimates[1] of the mean and standard deviation of the two velocities, as well as their correlation coefficient, for the stars in each shell. I then fit a smoothing spline to each of the parameters independently as a function of distance (right panels of Figure 2). The number of degrees of freedom of these fits is an increasing function of the number of shells, varying from 3 to 10 with a median of 8. Beyond the outermost radius the Gaussian parameters are kept constant at the value at the outermost radius. As the fit for the standard deviations could go to very small or even negative values, I force the fit of the two standard deviations never to drop below $2\,\mathrm{km\,s^{-1}}$.

The results of this process for a relatively sparse field at high latitude, where some shells are merged, are shown in Figure 2. In this particular example the variation of the correlation with distance is overfit; in other HEALpixels there is no clear correlation. For this reason I do not use these fits and set the correlation to be zero for all distances for all HEALpixels; a conser-

---

[1] I use the median and half the difference of the 84th and 16th percentiles of the distribution.
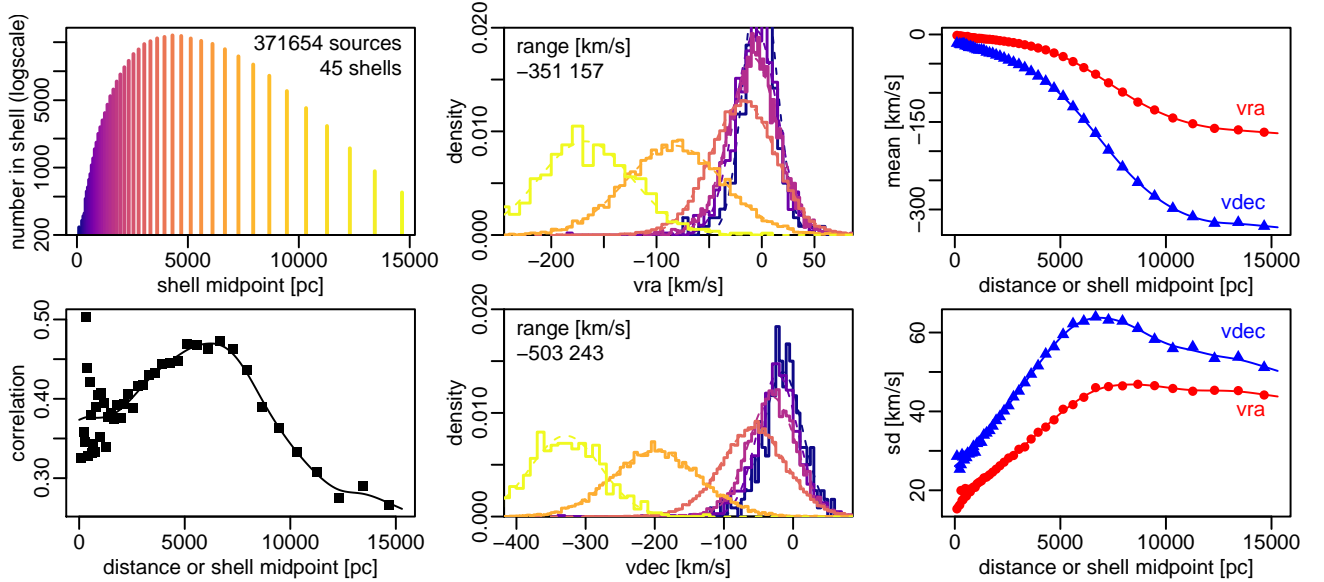
Figure 3. As Figure 2 but now for HEALpixel 7593 at HEALpixel 7593 at $(l, b) = (29.0°, 7.7°)$.

vative choice, making the prior less informative than it otherwise would be. Figure 3 summarizes the prior for a HEALpixel at a lower latitude towards the Galactic bulge, with many more stars where no shells are merged. Note the large change in the mean velocities for the most distant shells.

When building this velocity prior I excluded stars from the mock catalogue that are fainter than the expected Gaia magnitude limit for that HEALpixel, as was done for the distance prior in Bailer-Jones et al. (2021) (see section 1 of that reference).

I experimented with making the velocity prior a function of magnitude or colour. For magnitude this accounted for little additional variance. Introducing a dependence on colour was much harder, because some HEALpixels show gaps in their colour distribution, which would necessitate additional modelling assumptions to get a continuous prior.

Figure 4 shows how the magnitude of the mean of the velocity prior varies over the sky for several distance slices. The velocities of nearby stars are small, but increase at larger distances as we probe the faster orbits in the central regions of the Galaxy (note the change in colour scale across the slices). The low velocity regions near $(l, b) = (64°, 16°)$ and $(l, b) = (244°, −16°)$ in the 0 kpc distance slice are the approximate solar apex and antapex directions (respectively), the direction in which the Sun is moving through the Galaxy.

The inference (equation 3) uses the prior on the two components of the velocity, not the total transverse velocity. Plots like Figure 4 for the individual components are less astrophysically informative, however, because

they show features of the (equatorial) coordinate system used. An artefact of this is nonetheless visible at the south equatorial pole $(l, b) = (302.9°, −27.1°)$ in some of the distance slices in Figure 4. This is a projection effect of the velocity vector onto the rapidly changing $v_{\alpha*}$ and v_δ basis vectors at the poles. In principle we should not see it in the total transverse velocity, but because the prior is computed over the finite area of the HEALpixel, the average shows large jumps. The same effect occurs at the north equatorial pole, $(l, b) = (122.9°, +27.1°)$, but is barely visible. Using a smaller HEALpixel would mitigate this effect, but create others, and is a limitation of a discrete prior.

Figure 5 shows how the square root of the sum of squares of the standard deviations of the velocity prior varies. This too is not directly used in the inference, but it gives an idea of the width of the velocity prior, which influences how precisely distance can be constrained for a given proper motion (discussed in appendix A). The velocity constraint tends to be stronger at lower latitudes, as there is less dispersion about the mean rotational velocity of disk stars at a given distance and longitude. At higher latitudes the dispersion increases significantly because the more isotropic orbits of halo stars dominate the tails of the samples. While not immediately visible due to the variable colour scale, the velocity constraint is generally weaker (larger standard deviation) at larger distances.
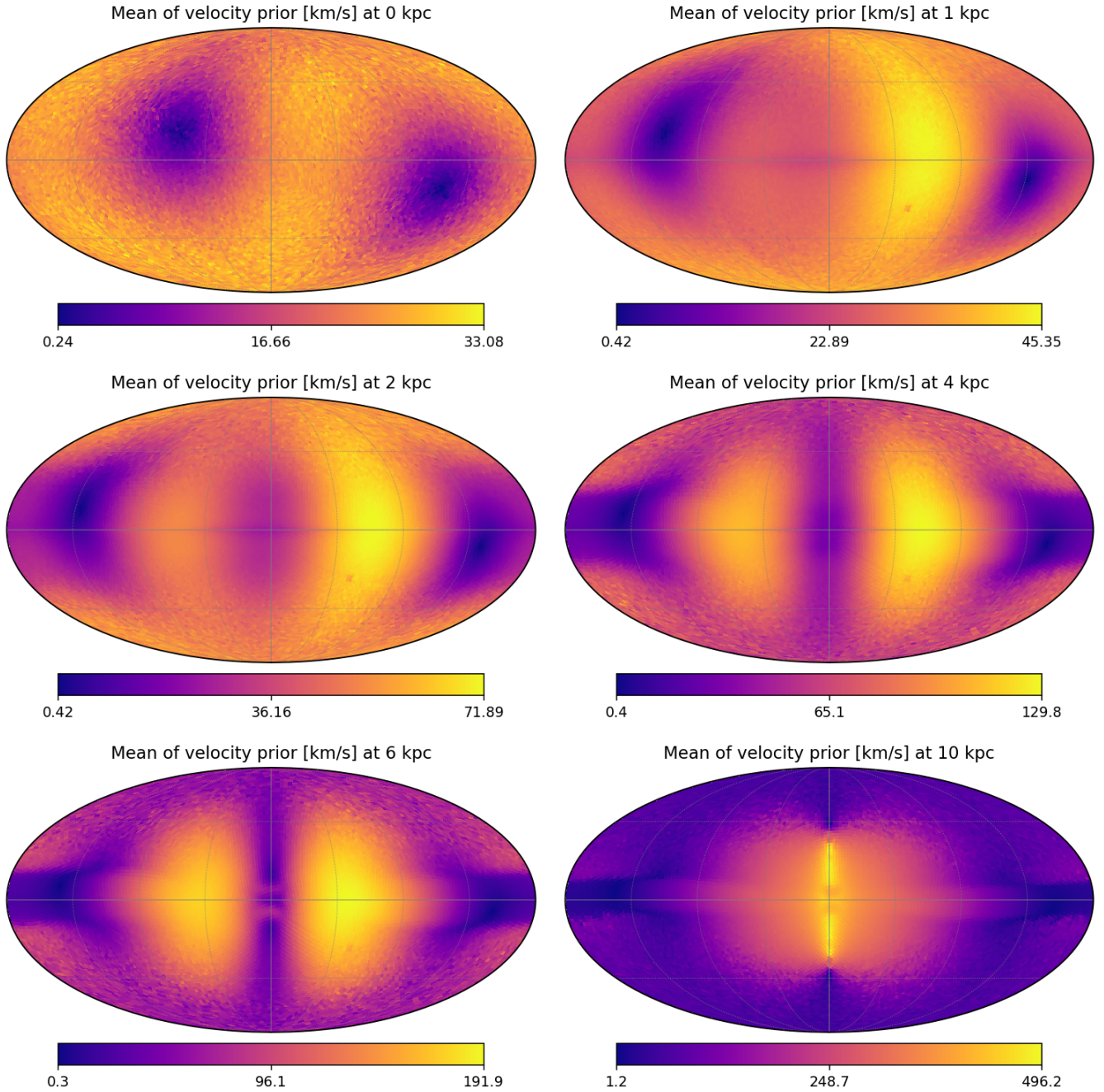
Figure 4. Variation of the mean of the velocity prior over the Galactic sky at six different distances. The quantity shown is $\sqrt{v_{\alpha*}^2 + v_\delta^2}$, where $v_{\alpha*}$ and v_δ are the means of the two components of the velocity prior at the specified distance in each HEALpixel. In each distance slice the colour bar covers the full range of velocities plotted, and is different for each slice.

2.3. *Posterior sampling*

I sample the 3D posterior (equation 3) for each star with an MCMC method (Foreman-Mackey et al. 2013).[2]

---

[2] As the posterior is Gaussian in velocity for a given distance, one could instead sample over distance and then compute the marginal posteriors for the velocity.

The distance is generally strongly correlated with the velocity components, because for a given proper motion an increase in the distance can be compensated for by a decrease in the velocity. Such correlations did not pose a problem for the sampler, and estimating instead the true angular velocities – which would not be correlated – did not confer particular advantages. After some experimentation I found that 60 burn-in iterations followed by
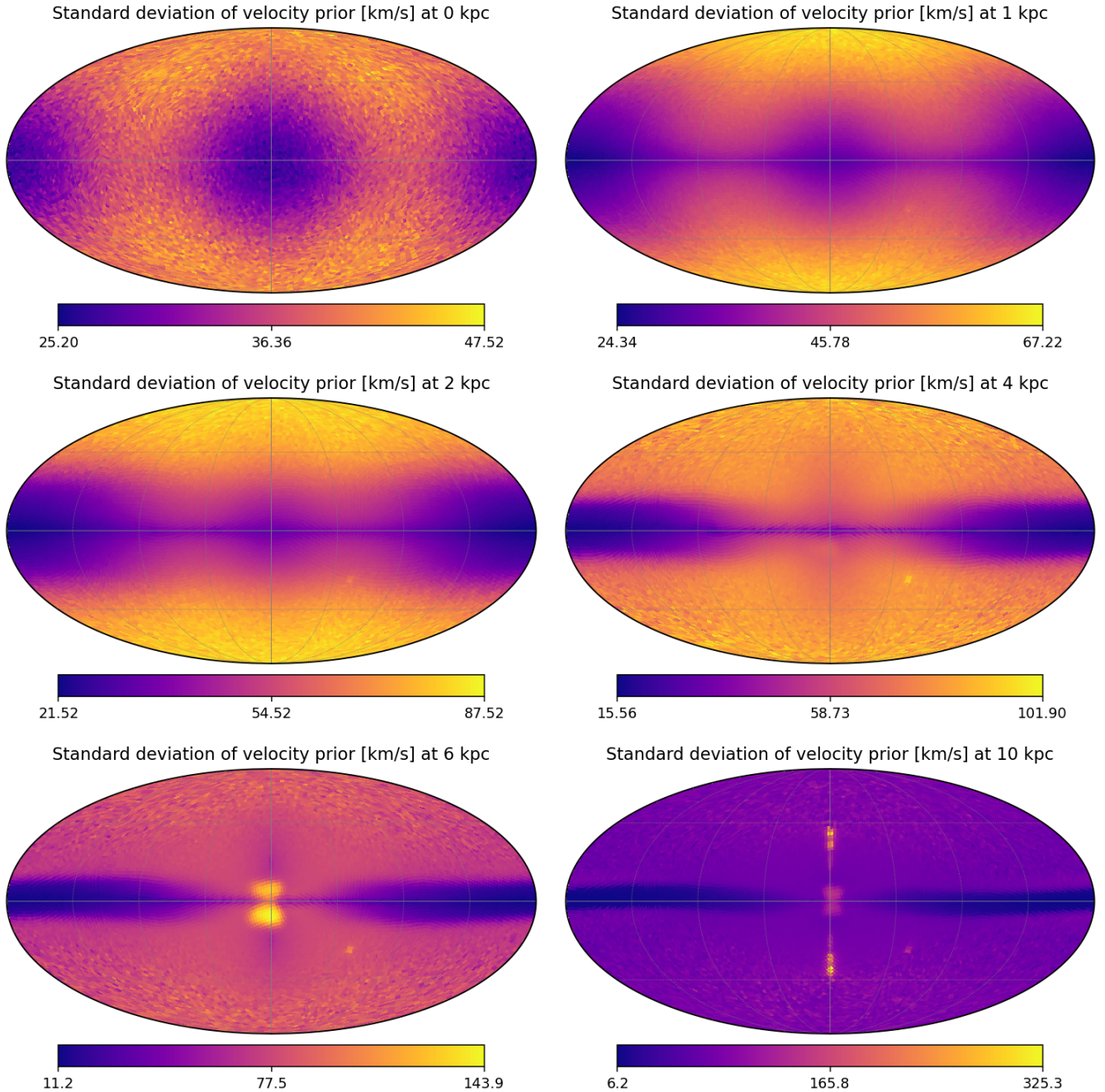
Figure 5. As Figure 4, but now showing $\sqrt{\sigma_{\alpha*}^2 + \sigma_{\delta}^2}$, where $\sigma_{\alpha*}$ and σ_{δ} are the standard deviations of the two components of the velocity prior.

200 sampling iterations with 20 walkers gave good convergence. The acceptance rate was around 0.6, and the sample chains were thinned before parameters were estimated. For each parameter I estimate the 16th, 50th, and 84th percentiles, the median providing the parameter estimate and the other two the lower and upper 1σ-like uncertainty estimates.

3. PERFORMANCE ON THE MOCK GDR3 CATALOGUE

To estimate the accuracy of the distance and velocity estimates, I apply the method to a noisy version of the mock Gaia catalogue. This is achieved by adding Gaussian random noise to the parallaxes and proper motions at the expected noise level in GDR3 (Rybizki et al. 2020). The noise-free version of this same catalogue was

used to build the priors, so this assessment shows what the method can achieve before being affected by any mismatch between the prior and reality.

I analyse the results over the whole sky using two samples. The *constant fraction sample* selects 0.8% of the stars at random from each HEALpixel to give a total of 10.3 million sources. Like the full Gaia sample, this is dominated by faint and low-accuracy parallax stars, many of which lie in the Galactic plane and bulge. While this is useful for examining trends in performance with distance and magnitude, it is not as useful for the examining the variation over the sky. For that purpose I use the *constant number sample*, which comprises approximately 900 sources selected at random per HEALpixel, for a total of about 11.5 million sources. Both samples include only stars that are brighter than the prior magnitude limit in each HEALpixel (section 2.2).

3.1. *Overall performance*

Figure 6 compares the estimated median distances and velocities against their true values for the constant fraction sample. The top left panel shows the geometric distance estimated from equation 2. The other panels show the kinegeometric distance and two velocity components $(v_{\alpha*}, v_\delta)$ from equation 3. The variations and differences are discussed in more detail below. Notable for now, though, is the smaller spread for kinegeometric distances compared to geometric ones at larger distances. From now on I will discuss and show results for the total kinegeometric velocity $\sqrt{v_{\alpha*}^2 + v_\delta^2}$ rather than its individual components. This will be compared with the total velocity derived from the geometric method, $kr_{\text{geo}}\sqrt{\mu_{\alpha*}^2 + \mu_\delta^2}$.

Performance statistics for the constant fraction sample are summarized in the top block of Table 1. For the distances I use the fractional residuals, (estimated-true)/true, because distance accuracy is a strong function of distance. For the total velocities I use the actual residuals, (estimated-true). We see that the residuals for the two methods are broadly similar, with a slightly heavier tail to negative residuals. For the distances, the fractional *bias* – median of (estimated-true)/true – is -0.0128 for the geometric distance and -0.0156 for the kinegeometric distance. For the velocities, the bias – median of (estimated-true) – is $-0.41\,\text{km s}^{-1}$ for the geometric method and $-0.88\,\text{km s}^{-1}$ for the kinegeometric one. Thus the biases are quite small on average.

The biases, which we can think of as the systematic deviation about the identity line in Figure 6, are part of the estimation error. The total estimation error I characterize by the median of the absolute values of the residuals, more specifically as median(|estimated-true|)/true) for

the distance and median(|estimated-true|) for the velocities. These robust versions of the error I refer to as the *scatter*. Note that it includes the bias. For the distances, the fractional scatters are 0.228 for geometric distances and 0.185 for kinegeometric distances. For the velocities, the scatters are $19.8\,\text{km s}^{-1}$ for the geometric method and $16.0\,\text{km s}^{-1}$ for the kinegeometric one. These are quite large, but it must be remembered that most of the stars in the sample are distant and faint.

The middle block of Table 1 shows the performance statistics on the constant number sample, where we see much smaller bias and scatter in all cases. This is because this sample has more stars at higher latitudes that are generally nearer. This is not as representative, so for global statistics I report the worse results on the constant number sample elsewhere in this paper. The final block is for GDR5, discussed in section 5.

The goal of the present paper is to introduce the kinegeometric estimates. The above statistics show that kinegeometric distances are overall more accurate than the geometric ones by a factor of only 1.25 (the ratio of the scatters). Within this they have a bias 1.2 times larger than the geometric estimates, although the average bias is only -1.5%. In the constant fraction sample, 58% of the stars have a kinegeometric error (scatter) that is smaller than the geometric error. 27% of stars have a kinegeometric error which is smaller than the geometric error by a factor of two or more (compared to 17% the other way around). The is no obvious, narrow part of parameter space where one estimate is always better than the other, but there are broad variations with distance and position, which we now explore.

3.2. *Performance as a function of distance*

Figure 7 shows how the bias and scatter of the kinegeometric distance estimate varies as a function of distance as a density plot, with the orange line indicating the median at each distance. We see a complex dependence on distance, in part because it is an average over sources over the whole sky. Compared to the geometric method (blue line), the kinegeometric bias and scatter are slightly smaller at most distances. The reason the geometric distance nonetheless showed a slightly smaller bias overall (Table 1) is because a large number of stars lie in the range where the geometric bias is smaller. Beyond about 7 kpc the kinegeometric distances are more accurate. This is the regime where the kinegeometric distances are most useful.

Figure 8 shows the performance as a function of parallax SNR (defined as the parallax divided by the parallax uncertainty). The performances of the two methods are similar in terms of bias, but the kinegeometric distances
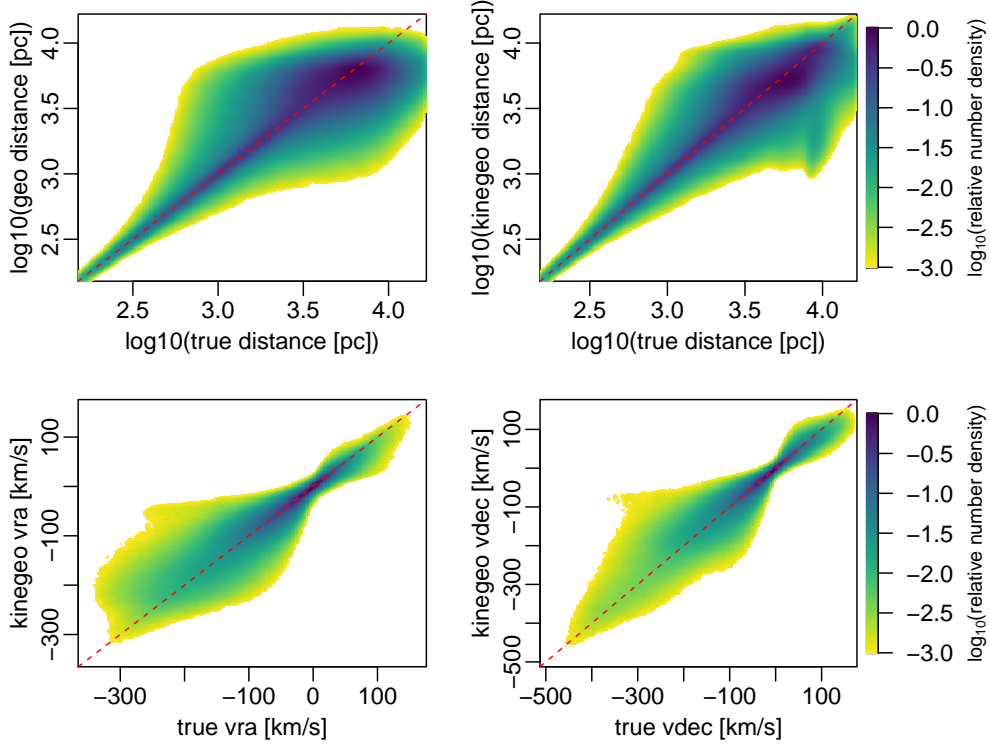
Figure 6. Estimated distances and velocities vs their true values for the constant fraction sample in the mock catalogue. The colour scale denotes the density of points on a logarithmic scale relative to the maximum in that panel, with densities less than a thousandth of the peak in white.

Table 1. Median performance for the two distance and two velocity estimates according to four different statistics, computed on the GDR3 mock catalogue using the two different samples (section 3) in the top and middle blocks. The bottom block shows the results for simulations at the higher precisions expected in GDR5, discussed in section 5.

|		geo distance	kinegeo distance	geo velocity	kinegeo velocity
constant fraction sample					
bias	est-true	$-20.9\,\mathrm{pc}$	$-29.9\,\mathrm{pc}$	$-0.410\,\mathrm{km\,s^{-1}}$	$-0.875\,\mathrm{km\,s^{-1}}$
scatter	| est-true |	$977\,\mathrm{pc}$	$779\,\mathrm{pc}$	$19.8\,\mathrm{km\,s^{-1}}$	$16.0\,\mathrm{km\,s^{-1}}$
fractional bias	(est-true)/true	-0.0128	-0.0156	-0.0098	-0.0178
fractional scatter	| (est-true)/true |	0.228	0.185	0.238	0.192
constant number sample					
bias	est-true	$-0.788\,\mathrm{pc}$	$-2.53\,\mathrm{pc}$	$-0.030\,\mathrm{km\,s^{-1}}$	$-0.192\,\mathrm{km\,s^{-1}}$
scatter	| est-true |	$199\,\mathrm{pc}$	$177\,\mathrm{pc}$	$6.14\,\mathrm{km\,s^{-1}}$	$5.68\,\mathrm{km\,s^{-1}}$
fractional bias	(est-true)/true	-0.0018	-0.0047	-0.0009	-0.0055
fractional scatter	| (est-true)/true |	0.125	0.112	0.128	0.114
constant fraction sample	for expected GDR5 astrometric precisions				
bias	est-true	$-10.1\,\mathrm{pc}$	$-18.1\,\mathrm{pc}$	$-0.229\,\mathrm{km\,s^{-1}}$	$-0.461\,\mathrm{km\,s^{-1}}$
scatter	| est-true |	$722\,\mathrm{pc}$	$595\,\mathrm{pc}$	$14.7\,\mathrm{km\,s^{-1}}$	$11.6\,\mathrm{km\,s^{-1}}$
fractional bias	(est-true)/true	-0.0072	-0.0105	-0.0057	-0.0106
fractional scatter	| (est-true)/true |	0.168	0.139	0.178	0.139
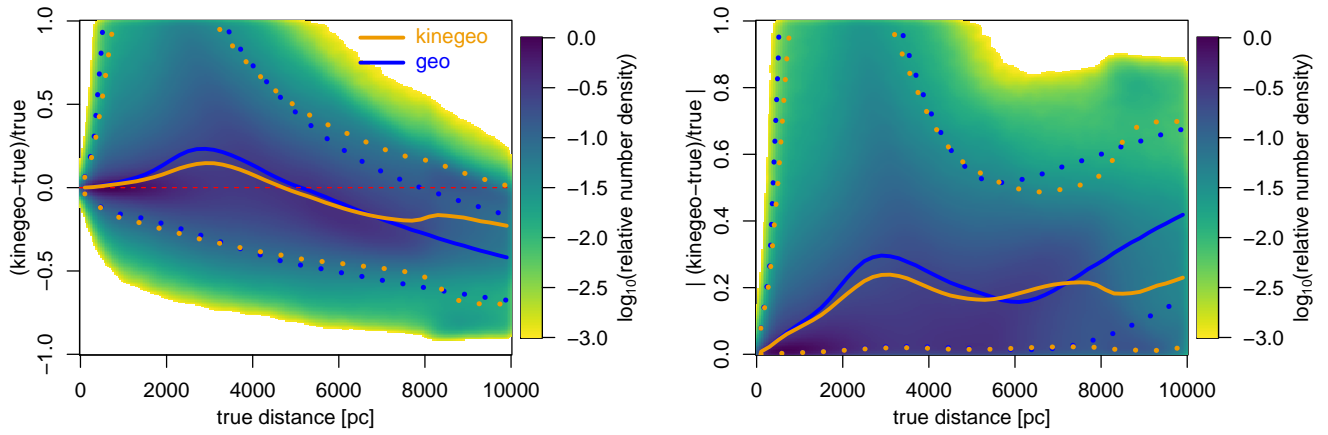
Figure 7. Performance of the kinegeometric distance estimate (median of the posterior) as a function of distance for the constant fraction sample in the mock catalogue. The left panel shows the fractional bias in the estimates, the right panel the fractional scatter. The colour scale shows the density of stars on a log scale. The solid orange line shows the median bias or scatter at each distance, the dotted lines are the 5th and 95th percentiles. The blue lines show the same metrics for the geometric distances for comparison.
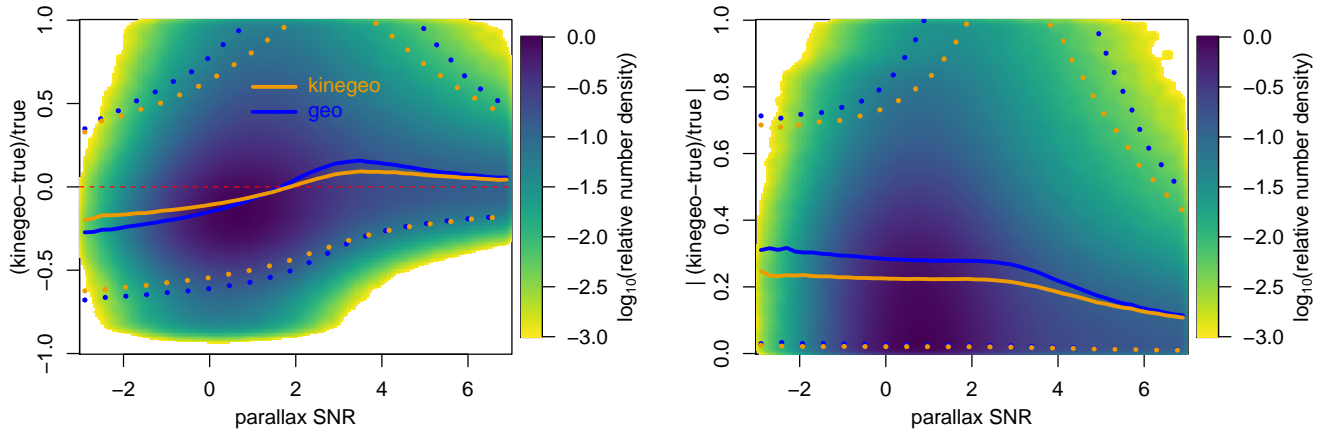
Figure 8. As Figure 7, but now as a function of parallax SNR.

have lower scatter for low parallax SNR. This is precisely the region where the proper motions are helping, because even though the proper motion SNRs are generally lower too, they are still high enough to provide additional information on distance. At higher parallax SNR the differences are small because the parallax dominates, and the priors are less important.

Figure 9 compares the bias and scatter of the two total velocity estimates as a function of distance. Again we see a large range of these statistics at a given distance, as evidence by the dotted lines in these plots. The velocity estimates have similar performances (in both bias and scatter) for stars with true distances less than about 7 kpc. At larger distance the kinegeometric velocity estimates are more accurate, due to the assistance from the proper motions and velocity prior.

3.3. *Performance as a function of direction*

We turn look at how the performance varies over the sky using the constant number sample.

Figure 10 shows the fractional bias and scatter, by computing the statistic for each star and then taking the median over each HEALpixel. Looking first at the upper panels (the bias), we see that both methods tend to underestimate distances in lines-of-sight in the disk and bulge. This is because these directions are dominated by distant stars (see Figure 1) which generally have low parallax SNRs, which in turn introduces a bias (explained in section B). The kinegeometric estimates show larger negative biases by up to an average of 13% in two regions above and below the Galactic centre. This is related to the large changes in the velocity prior beyond a few kpc (see Figure 4). The geometric distance estimates are less affected.

The scatter plots (bottom row of Figure 10) show a larger scatter in both distance estimates at low latitudes. This is a consequence of most stars being at large true distances, so have low SNR parallaxes. The kinegeometric distances show a smaller scatter though, because the proper motions improve the distance accuracy primarily at low parallax SNR (as already discussed in the context of Figure 8).

Figure 11 shows the differences between the kinegeometric and geometric distance estimates over the sky (computing the median per HEALpixel of the differences for each star). At higher Galactic latitudes the estimates are very similar on average. In the central regions of the Galaxy the kinegeometric distances are smaller, by up to 11% on average within a HEALpixel. In the rest of the Galactic disk the kinegeometric distances are larger by up to 8% on average.

Figure 12 shows how the performances of the two velocity estimates vary over the sky. The bias patterns are broadly similar to what we saw with the distances, especially for the kinegeometric estimates, e.g. the blue blobs at $l \simeq 0°$ mentioned above. The bias map for the geometric velocities, however, shows two regions in the disk with a small but significant overestimate, at around $l = -155°$ and $l = 150°$ (the "red cheeks" in the top-left panel of Figure 12). They are slightly displaced from the Galactic plane, so are presumably related to the warp of the disk in the Galaxy model on which the mock catalogue is based. They do not appear in the bias map for kinegeometric distances (top-right panel of Figure 12). Some structures in these regions are seen in the nearer slices of the velocity prior (Figure 4), but not in the distance prior (Figure 1), suggesting that not accounting for these in the geometric prior is problematic. Naturally, if these features do not exist in the real Galaxy, the (unknown) biases in the inference in GDR3 could be different.

The lower panels of Figure 12 show the scatter in the two velocity estimates, which are similar in pattern, but slightly larger for geometric velocities. Figure 13 shows the average differences per HEALpixel between the two velocity estimates.

3.4. *Uncertainty estimates*

From the posterior PDF for each star I compute the 16th and 84th percentiles for each parameter, then take their difference from the median to obtain lower and upper 1σ-like uncertainty estimates. Half their difference gives a single uncertainty estimate for the parameter. If the residuals (median minus true) have an unbiased Gaussian distribution, and if the uncertainty is a statistically correct estimate of the residual, then the ratio of the residual to the uncertainty – the normalized residuals – should have a unit Gaussian distribution. These are shown in Figure 14. All four distributions have a mean close to zero and a standard deviation close to unity, although the distributions for both distance measures show a slight negative skew. This is not surprising because we know there is a negative distance bias for distant, low parallax SNR stars (see appendix B). But overall this shows that the uncertainty estimates are good statistical estimates of the true error.

4. RESULTS ON GDR3

1.46 bilion stars in GDR3 have parallaxes and proper motions. As with the mock catalogue, I apply the geometric and kinegeometric inference methods to two randomly-selected subsets: a constant number sample with 0.8% of all sources (10.3 million sources),
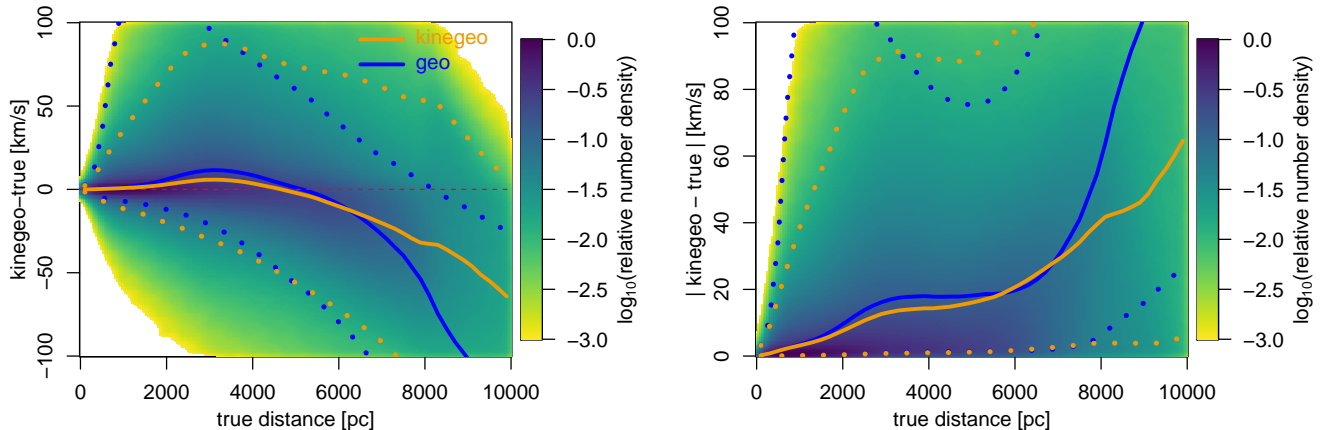
Figure 9. Performance of the kinegeometric velocity estimate (median of the posterior) as a function of distance for the constant fraction sample in the mock catalogue. The left panel shows the bias in the estimates, the right panel the scatter. The colour scale shows the density of stars on a log scale. The solid orange line shows the median bias or scatter at each distance, the dotted lines the 5th and 95th percentiles. The blue lines show the same metrics for the geometric method (geometric distance times proper motion).

and a constant fraction sample with 900 sources per HEALpixel (10.8 million sources), in both cases restricted to sources brighter than the prior magnitude limit in each HEALpixel (Bailer-Jones et al. 2021, section 1) to enable comparison to the estimates on the mock catalogue in the previous section.

4.1. *Distances*

Figure 15 shows the median distance per HEALpixel. The median distance increases to lower Galactic latitudes in both methods, because of the higher density of stars at larger distances within the disk. The median distance is smaller in the Galactic plane, however, because Gaia's depth is limited by strong interstellar dust extinction. Many of the patterns we see close to the plane are due to dust density variations. At high Galactic latitudes the depth through the disk is small, and although distant stars in the halo are visible, they are rare, so the median remains relatively small. The bulge is quite prominent in both distance estimates as a larger, approximately circular region of more distant stars. Close to the Galactic centre we see dust-free lines-of-sight that extend the median distance to over 7 kpc. Neither the mock catalogue as we use it nor the prior contains star clusters, so cluster distances will not be correct. This is most noticeable for stars in the Large and Small Magellanic Clouds, which we see have severely underestimated distances.

The differences between the two distance estimates are shown in the bottom panel of Figure 15 (median of differences per star). The largest differences are in parts of the Galactic plane, parts of the bulge, and in several star clusters. Most prominent is that the ki-

negeometric distance is smaller in the bulge above and below the Galactic centre. This is a result of the measured proper motions combined with the velocities of the Galaxy prior. In parts of the Galactic plane the kinegeometric distances are larger, in others the differences are negligible, and in the northern part of the disk near the anticentre the kinegeometric distances are smaller. These patterns are presumably related to the Galactic bar and warp, which unsurprisingly differ between GDR3 and the Galactic model used to make the prior.

The larger kinegeometric distances in the upper and lower parts of the bulge was also seen in the results on the noisy mock data in Figure 11, although it was less prominent there and also extended through the Galactic disk. We should not over interpret differences between Figures 11 and 15, however: they can occur for many reasons, including differences in the underlying stellar spatial and velocity distributions or in the extinctions between the mock catalogue and real Galaxy. The larger colour bar scale on the difference plot for the GDR3 results shows that kinegeometric and geometric distances differ much more in the real inference than they did in the mock inference, no doubt a result of these differences.

Figure 16 compares the two distance estimates as a function of geometric distance. The density plot and the quantiles (dashed lines) show that there is a large spread, but the median trend (solid orange line) is for the kinegeometric distances to be on average smaller than the geometric ones beyond 6 kpc, by 20%
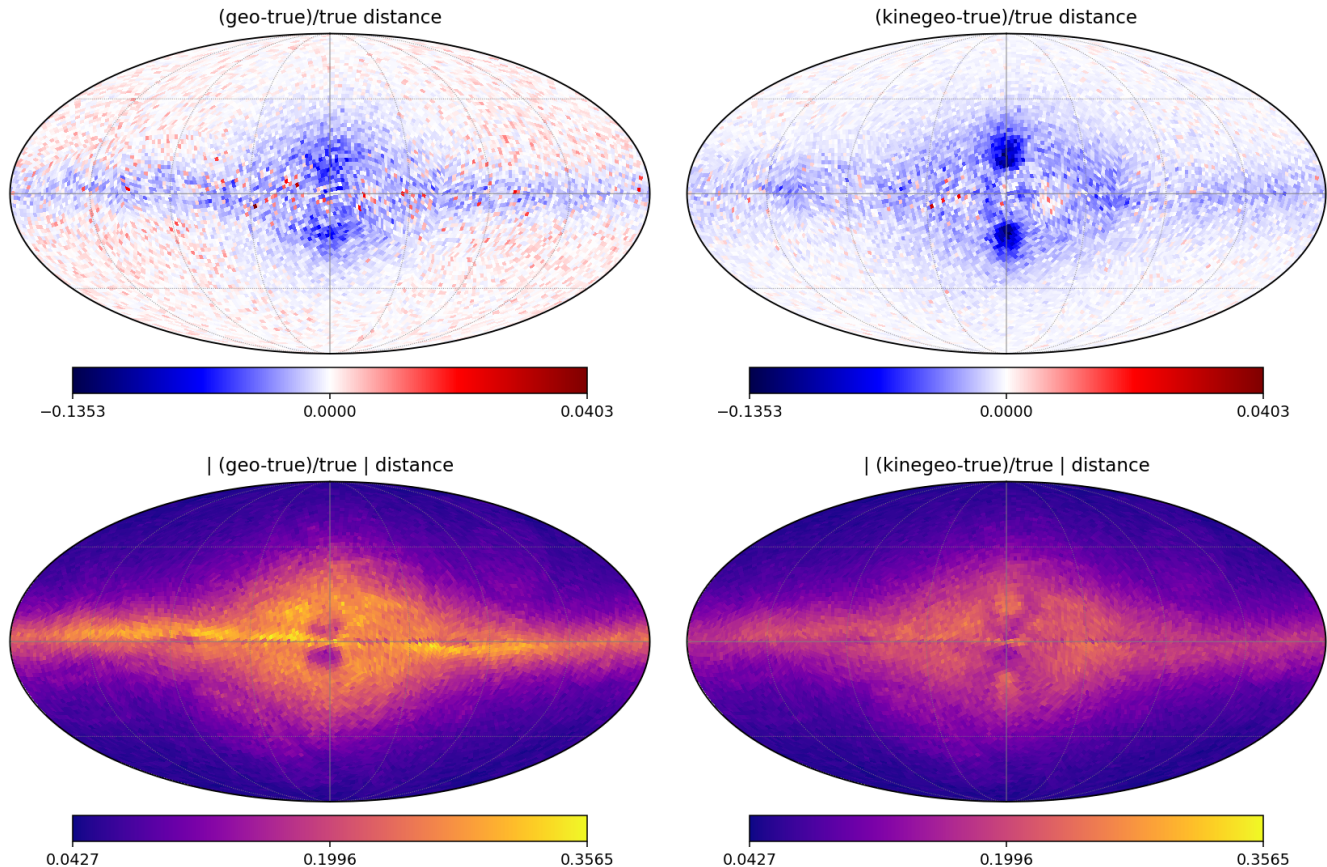
Figure 10. Fractional distance residuals – (estimated-true)/true – per HEALpixel for the constant number sample in the mock catalogue. The top row shows the median fractional residual (the bias); the bottom row shows the median absolute fractional residual (the scatter). The left column is for geometric estimates, the right column is for kinegeometric estimates. The colour bars span the full common range in each of the top and bottom rows. The top row uses a bilinear colour bar (separate scales for negative and positive values): The dominance of red HEALpixels over blue ones at high latitudes is because the positive red scale extends to smaller values than the negative blue scale; the absolute biases are more or less the same overall.

at 10 kpc.[3] This is not true in all lines-of-sight, however, such as the Galactic centre and parts of the disk, as we saw in Figure 15. When we consider just the low parallax SNR stars (< 3), we see a large difference in the distance estimates on average also for nearby stars (the dashed pink line in Figure 16).

We saw in section 3.4 (Figure 14) that the width of the posterior (the precision) was a good statistical estimate of the actual error. We can therefore compare the posterior width (the difference between the 84th and 16th percentiles) for the two types of distance estimate to predict which is more accurate. Using the constant fraction sample, we find that for 62% of the stars, the

kinegeometric posterior is narrower than the geometric one. 15% of stars have a kinegeometric posterior that is less than half the width of the geometric one, compared to just 4% the other way around. This suggests kinegeometric distances will be more accurate for a significant number of stars. Figure 17 plots the ratio of these posterior widths over the sky, taking the median over all stars in each HEALpixel. We see that in almost all parts of the sky – but particularly in the bulge and disk – the kinegeometric distances are more precise (but remembering that each HEALpixel is an average over a large range of distances). We also find that kinegeometric distances tend to be more precise for lower parallax SNRs, as we would expect.

4.2. Velocities

Figure 18 shows the median velocity per HEALpixel inferred by both the geometric method (geometric distance times proper motion) and the kinegeometric

---

[3] One might read the left panel of Figure 7 to suggest we had the opposite in the mock data, but that plot shows the biases against the *true* distance. If we replace true with geometric, in then looks very similar to Figure 16.
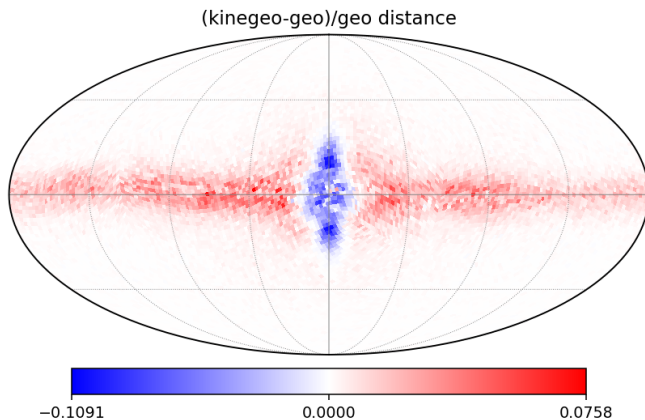
(kinegeo-geo)/geo distance

−0.1091   0.0000   0.0758

Figure 11. Median fractional differences between the kinegeometric and geometric distances per HEALpixel (computed on the individual stars), for the constant number sample in the mock catalogue. The colour bar has white at zero and diverges on different linear scales to the maximum negative and positive differences.

method. These are transverse velocities relative to the solar system barycentre. The largest velocities occur within about 90° in longitude of the Galactic centre, particularly at low latitudes. This is where we see mostly disk stars on smaller radius orbits than the Sun's. The exception is in the Galactic plane, where dust extinction means Gaia only sees nearby stars that are orbiting the Galaxy with low velocities relative to the Sun. At larger longitudes we see the outer part of the disk where, due to the flattening of the Galactic rotation curve, stars are orbiting at speeds more similar to the Sun's.

The differences between the two velocity estimates are shown in the bottom panel of Figure 18. This is broadly similar to the difference plot obtained from the mock catalogue results in Figure 13, although the scales are different. The main difference (other than the absence of star clusters in the mock results), is around the Galactic centre: whereas for mock the geometric velocities were several km s^{-1} larger on average, in GDR3 the kinegeometric velocities are slightly larger.

5. EXPECTED PERFORMANCE WITH THE FINAL GAIA DATA RELEASE (GDR5) AND THE LIMITS OF PROPER MOTIONS

The distance accuracy that this method can attain is determined by the accuracies of the parallaxes and proper motions, as well as the width of the distance and velocity priors. The accuracy of the data will improve in later Gaia data releases. If t is the timespan of observations, then whereas the parallax accuracy im-

proves as $t^{1/2}$ on account of the increase in the number of observations, the proper motion accuracy improves as $t^{3/2}$ due to the additional increase in the observational baseline. This assumes systematic errors do not dominate, which is the case for most stars. GDR3 is based on 34 months of data, whereas the final release, GDR5, will be based on about 10.5 years (126 months) of data. Using these factors, the parallax and proper motion accuracies should therefore increase by factors of 1.9 and 7.1 respectively between GDR3 and GDR5. This suggests that while both the kinegeometric and geometric distance estimates should be more accurate with GDR5 data, the improvement for the kinegeometric ones might be larger.

I tested this expectation using the mock catalogue with similar simulations as before (section 3), but now scaling down the parallax and proper motion uncertaintines by the above factors (1.9 and 7.1 respectively). The results for the constant fraction sample are shown in the bottom block of Table 1. Looking first at the geometric distances – which depend on the parallaxes but not the proper motions – we see that the fractional bias is decreased from GDR3 to GDR5 by a factor of 1.8 and the fractional scatter is decreased by a factor of 1.35. These improvements are not as large as the improvement in the parallaxes. This is expected, because the prior also plays a role: The distance accuracies for the numerous stars with very low parallax SNRs ($\lesssim 1$) are only weakly dependent on their parallax SNRs because their distances are prior-dominated. For such stars, doubling the parallax SNR hardly changes their distance estimates and therefore hardly improves their accuracies.[4]

For the kinegeometric distances, the fractional bias and scatter are decreased by factors of 1.55 and 1.35 respectively. Contrary to possible expectations, the kinegeometric distances do not improve more than the geometric distances: On average, at least, the larger improvement in the proper motions over the parallaxes has not helped. The reason is that the proper motion only provides a distance estimate when combined with an assumed velocity distribution (the prior) via equation 1. Once the proper motions are sufficiently precise such that the width of this velocity distribution is the limiting factor in determining the distance precision, then improving the proper motions further will not help. This is explained in appendix A. It appears that, on average at least, this limit has already been achieved in GDR3.

---

[4] Even with weaker priors, such as uniform in distance (or log distance), doubling the parallax SNR does not double the accuracy of the distance estimate, because of the nonlinear transformation between parallax and distance.
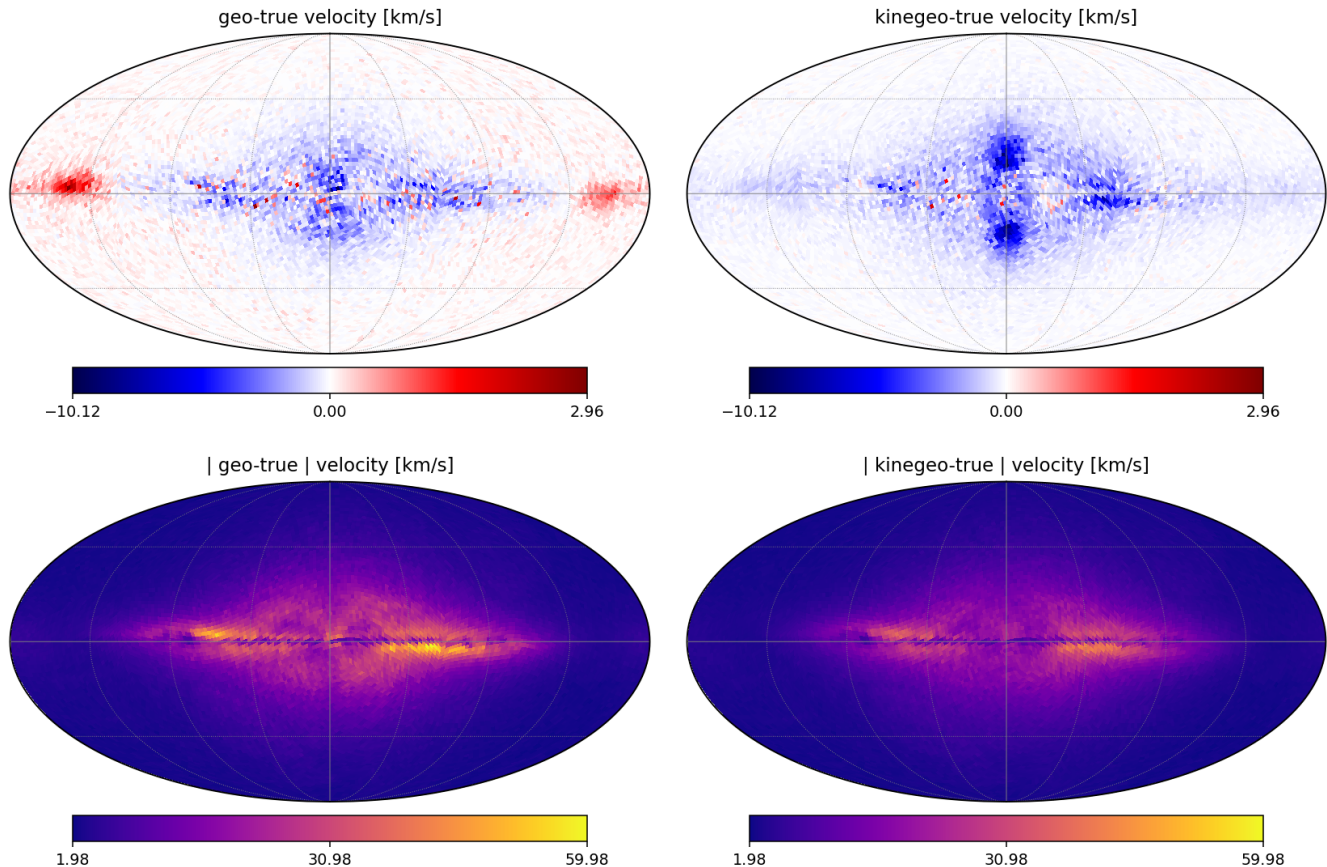
Figure 12. Velocity residuals – (estimated-true) – per HEALpixel for the constant number sample in the mock catalogue. The top row shows the median residual (the bias); the bottom row shows the median absolute residual (the scatter). The left column is for geometric method (geometric distance times proper motion), the right column is for kinegeometric estimates. The colour bars span the full common range in each of the top and bottom rows. The top row uses a bilinear colour bar (separate scales for negative and positive values).

The variation of the distance accuracy (bias and scatter) with distance for GDR5 is shown as the dashed lines in Figure 19. The profiles are broadly similar to GDR3 (solid lines), but we see that the improvements in the accuracy vary with distance. Below 2 kpc the scatter in both the geometric and the kinegeometric distances is reduced by a factor of nearly two, whereas at around 5 kpc there is little improvement.

For the velocities the situation is similar (Table 1). The scatter in the geometric velocities is expected to decrease by a factor of 1.35 times between GDR3 and GDR5, and by a factor of 1.4 for kinegeometric velocities. We might have expected much more improvement in both, because the proper motions improve by a factor of 7.1. But the mapping from proper motion to velocity depends also on the estimated distance, and the improvement in this, which is coming mostly from the parallaxes, is also a limiting factor in the improvement in the velocities.

I conclude that, in GDR5, kinegeometric distances and velocities should be about 1.35 times more accurate than they are in GDR3, when averaged over the entire Gaia sample in GDR3. These will still be only about 1.25 times more accurate that geometric distances and velocities. The limited improvement is because the velocity priors are relatively broad.

While it may be worth using proper motions in GDR5 to improve the distance estimates in some regions, this comes at the cost of building in assumptions on velocities. If we are happy with this, then a better accuracy should be obtained by combining parallaxes and proper motions also with colours, magnitudes, and even spectra.

6. SUMMARY AND CONCLUSIONS

The kinegeometric method introduced in this paper estimates distances and transverse velocities together in a 3D posterior PDF using the parallaxes and proper motions. It uses a direction-dependence distance prior
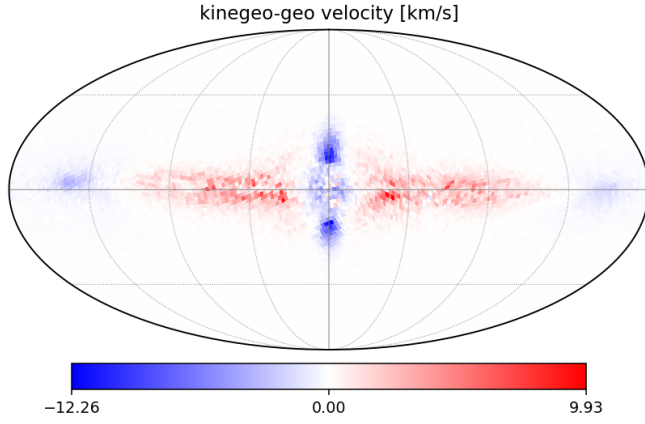
kinegeo-geo velocity [km/s]

−12.26   0.00   9.93

Figure 13. Median differences between the kinegeometric and geometric velocities per HEALpixel (computed on the individual stars), for the constant number sample in the mock catalogue. The colour bar has white at zero and diverges on different linear scales to the maximum negative and positive differences.
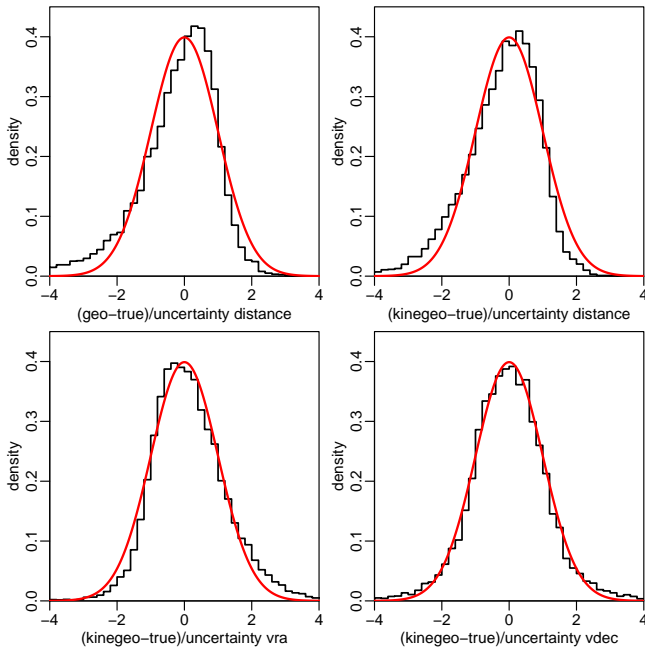

Figure 14. Normalized residuals for the geometric posterior distance (top left) and for the three parameters of the kinegeometric posterior (other panels), computed for all sources in mock HEALpixel 6200. The smooth red lines show the unit Gaussian for comparison.

and a direction- and distance-dependent velocity prior. Geometric distances, in contrast, are obtained from just the parallaxes and the distance prior.

The performance was assessed on a random subset of the mock GDR3 catalogue. This showed that on aver-

geo distance [pc]

1000   4188   7375

kinegeo distance [pc]

1000   4188   7375

(kinegeo-geo)/geo distance

−0.2468   0.0000   0.2834
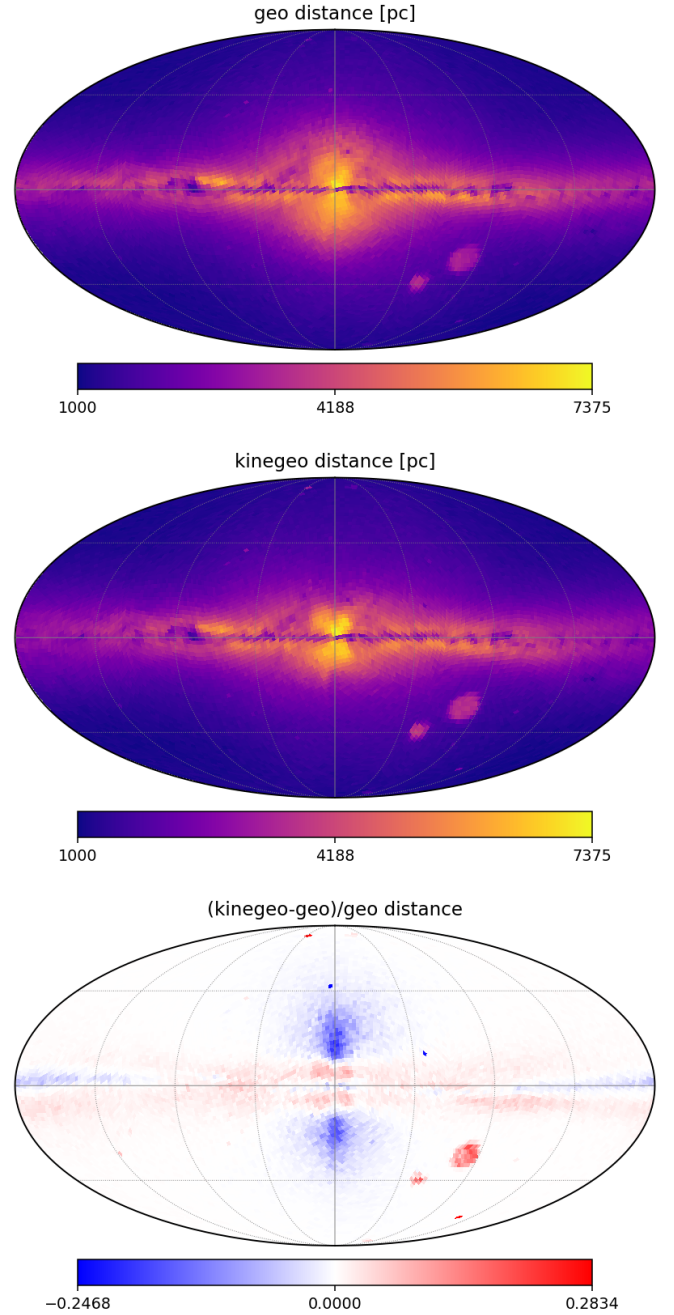
Figure 15. Median distance per HEALpixel for the constant number sample in GDR3: Geometric (top) and kinegeometric (middle). The linear colour bar spans the full common range. The bottom panel shows the median fractional differences between these (computed on the individual stars). It uses a bilinear colour bar with white at zero and diverges on different linear scales to the maximum negative and positive differences.
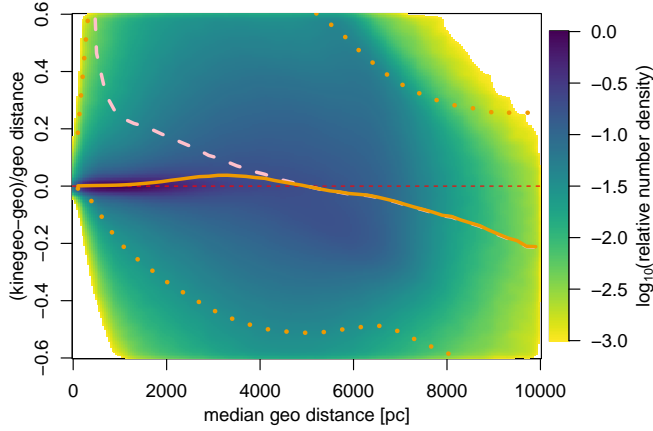
Figure 16. Variation of the fractional difference between the kinegeometric and geometric distances with geometric distance for the constant fraction sample in GDR3. The colour scale shows the density of stars on a log scale. The solid orange line shows the median fractional difference, the dotted lines the 5th and 95th percentiles. The dashed pink line shows the median fractional difference for just the subset with parallax SNR less than three.
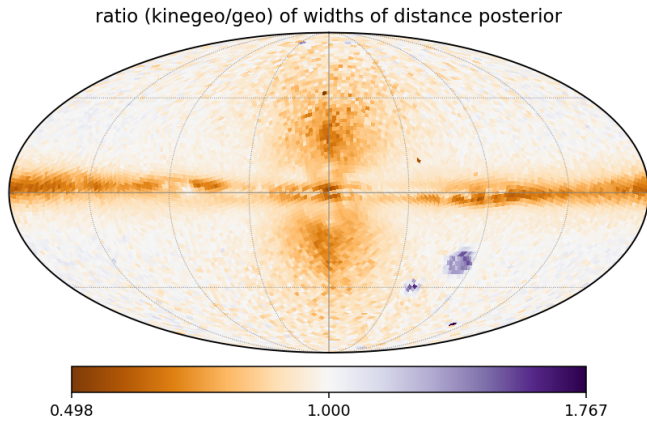

Figure 17. Median per HEALPixel of the ratio (kinegeo/geo) of the posterior widths for the constant fraction sample in GDR3. Numbers less than 1 (in orange) show where the width of the kinegeometric distance posterior is narrower than the width of the geometric distance posterior, and are therefore statistically likely to be more accurate. We see that this is the case in most parts of the sky, but particularly in the disk and bulge. The bilinear colour bar with white at unity and diverges on different linear scales to the minimum and maximum values.

age, kinegeometric distances are 1.25 times more accurate than geometric ones (Table 1). The median fractional distance residual (scatter) of the kinegeometric distances is 19% averaged over all sources; averaged over the sky it is 11%. The main improvement brought by
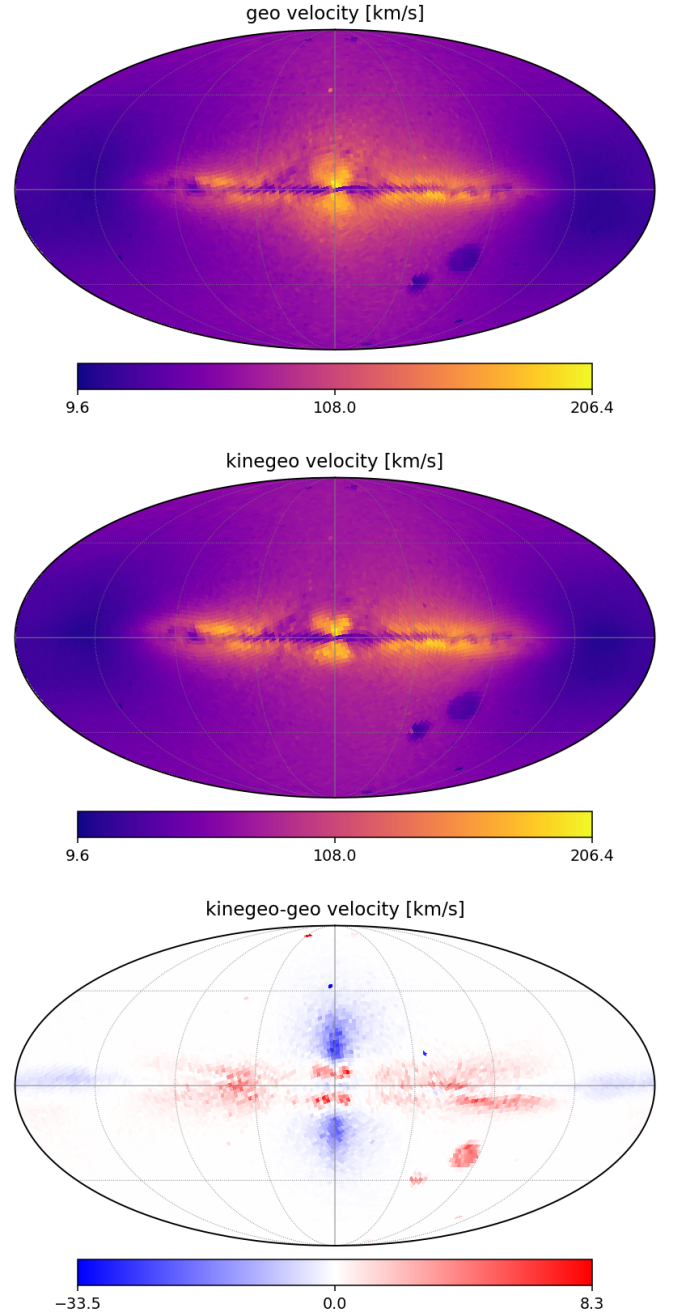

Figure 18. Median transverse velocity per HEALPixel for the constant number sample in GDR3: Geometric (top) and Kinegeometric (middle). The linear colour bar spans the full common range. The bottom panel shows the median differences between these (computed on the individual stars). It uses a bilinear colour bar with white at zero and diverges on different linear scales to the maximum negative and positive differences
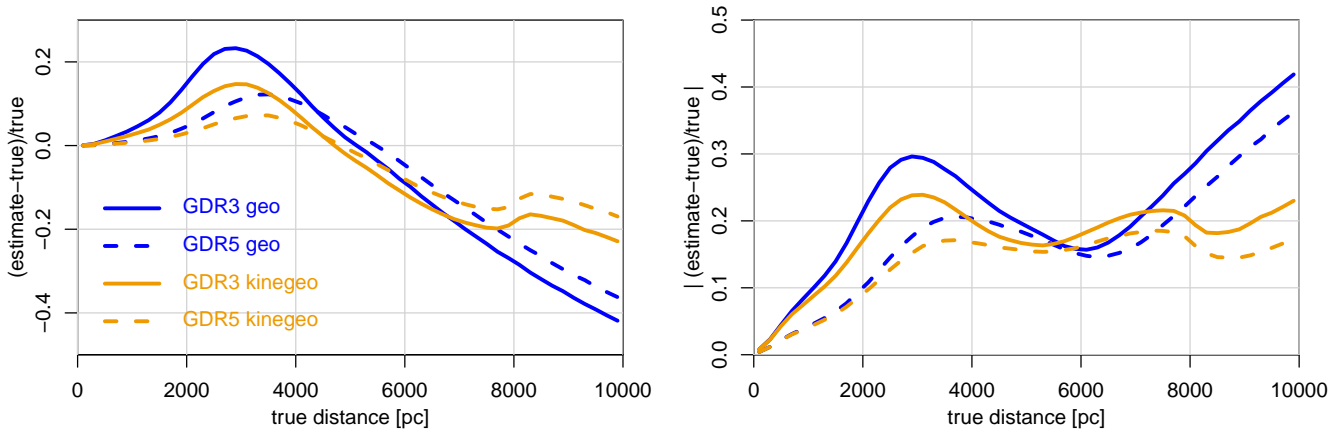
Figure 19. Median performance of the geometric (blue) and kinegeometric (orange) distance estimates as a function of distance for the constant fraction sample in the mock GDR3 (solid) and mock GDR5 (dashed) catalogues. The left panel shows the fractional bias in the estimates, the right panel the fractional scatter. The solid lines are the same as those in Figure 7.

the proper motions is where the parallax SNR is low ($\lesssim 4$; Figure 8). There is a considerable spread in performance with position in the Galaxy, however. Close to the Sun the kinegeometric distance errors are very small, then increase steadily to 20% at 2 kpc and beyond (Figure 7). Kinegeometric distances have a low bias at high Galactic latitudes, but up to several percent in the bulge and disk, with complex patterns emerging at low latitudes due to the stellar density and prior variations (Figure 10). The biggest advantage of kinegeometric distances over geometric ones is for stars beyond about 7 kpc from the Sun.

The kinegeometric method also provides transverse velocity estimates. These can be compared to "geometric" velocities obtained by multiplying the geometric distances by proper motion. Both approaches have similar overall performance for stars within 7 kpc, but beyond that the kinegeometric velocities are considerably more accurate, including a lower bias (Figure 9).

When applied to real GDR3 data we find that differences in the two distance estimates depend on direction and distance. Beyond several kpc, kinegeometric distances tend to be smaller on average (Figure 16), which in particular is reflected by two regions in the bulge (Figure 15). In the disk the picture is more complicated, a reflection of how the disk kinematics in the prior combine with the measured proper motions to inform the distance estimates. The precisions (widths) of the posteriors are a statistically good estimate of the accuracies (Figure 14), so we can use them to investigate the expected accuracy of the distance estimators in the absence of a ground truth. The kinegeometric distances are more precise than the geometric ones for 62% of stars, and (on average) over most of the sky (Figure 17).

Overall, we see only a modest average improvement in distance accuracy when including proper motions. This conclusion is based on the 34 months of Gaia data in GDR3. The SNR of the parallaxes and proper motions are expected to improve in future Gaia data releases, by factors of 1.9 and 7.1 respectively in the final release (GDR5) assuming this is based on 10.5 years of data. I find that this will improve the accuracy of the kinegeometric distances and velocities by a factor of 1.35. The geometric distances improve by a similar factor, meaning kinegeometric distances are still only about 1.25 times more accurate than geometric ones in GDR5. The reason for this limited improvement is that proper motions only constrain distances when combined with an expected velocity, which we introduce as a distribution of velocities. If this prior is broad, then even very precise proper motions do not constrain the distances much (see appendix A). This is very different from the situation with parallaxes, in which an increased accuracy of the parallax translates directly into an increased accuracy in the distance (provided the distance is not prior-dominated). Hence, for many stars in Gaia, the velocity distributions in the Galaxy are too broad for even high precision proper motions to provide strong constraints on distance.
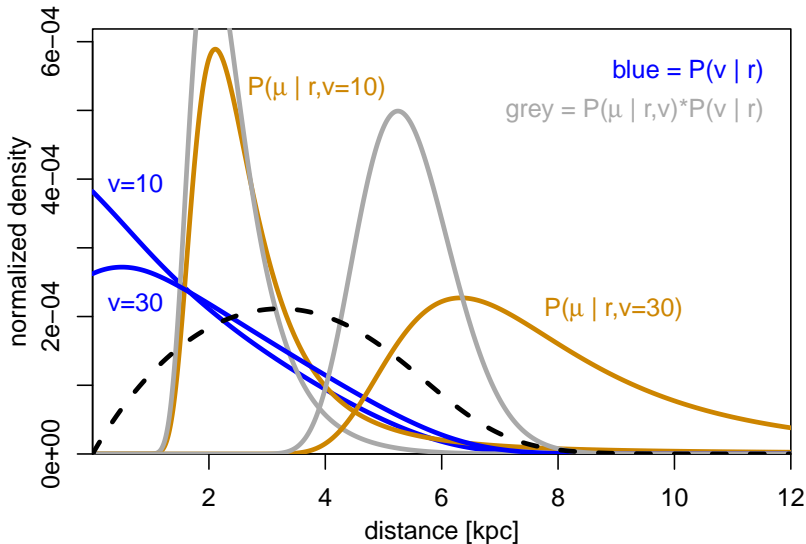
Figure 20. Demonstration of how the proper motion likelihood ($P(\mu \mid r, v)$, orange) combines with the velocity prior ($P(v \mid r)$, blue) to constrain the distance via equation A1, from a measured proper motion of $1 \pm 0.25\,\mathrm{mas\,yr^{-1}}$. This is shown for two different velocities, 10 and $30\,\mathrm{km\,s^{-1}}$. The grey curve is the product of the orange and blue curves. The black dashed line is the integral in equation A1 computed for a continuous range of velocities. None of these functions are probability density functions in r, but all have been normalized to unit area over the range plotted.

DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

Facility: Gaia

APPENDIX

A. HOW THE PROPER MOTION AND VELOCITY PRIOR CONSTRAIN DISTANCE

The kinegeometric posterior of equation 3 is a 3D distribution coupling distance and velocity. We can understand how the proper motion and velocity prior combine to produce distance estimates by considering a simpler version with just one velocity component and no parallax. In this case the posterior over distance is

$$P(r \mid \mu) = \int P(r, v \mid \mu)\, dv = \int \frac{P(\mu \mid r, v) P(r, v)}{P(\mu)}\, dv \propto P(r) \int P(\mu \mid r, v) P(v \mid r)\, dv . \tag{A1}$$

We see that the proper motion likelihood constrains the distance only through the integral of its product with the velocity prior. This is quite different from the parallax, where $P(r \mid \varpi) \propto P(r) P(\varpi \mid r)$ without any integral.[5] Consequently, even if we measure a very precise proper motion, such that the likelihood $P(\mu \mid r, v)$ for some v in equation A1 is narrow, if the corresponding velocity prior $P(v \mid r)$ is broader, the contribution to the posterior will also be broad.

This combination is demonstrated in Figure 20. I use the v_δ prior from HEALpixel 7593 (Figure 3) except that I take the absolute value of the mean of the velocity prior to more conveniently deal with positive values. The prior (blue curves) is shown for 10 and $30\,\mathrm{km\,s^{-1}}$. Both velocities show a preference for smaller distances. The two orange curves are the likelihoods corresponding to these same two velocities assuming a proper motion measurement of $1 \pm 0.25\,\mathrm{mas\,yr^{-1}}$. $1\,\mathrm{mas\,yr^{-1}}$ and $10\,\mathrm{km\,s^{-1}}$ corresponds to distance of $2.1\,\mathrm{kpc}$ (and $30\,\mathrm{km\,s^{-1}}$ to $6.3\,\mathrm{kpc}$; equation 1), but we see that this low SNR measurement does not constrain the distance much for either velocity. Each grey curve is the product of the prior and likelihood, and therefore a term under the integral in equation A1 for a given velocity.

If we repeat this for a continuous range of velocities, then the sum of all the resulting grey curves is the integral in equation A1, which is the unnormalized distance posterior (but omitting a distance prior). This is illustrated in the

---

[5] This follows by replacing μ with ϖ in equation A1 and then noting that the parallax likelihood is independent of v.
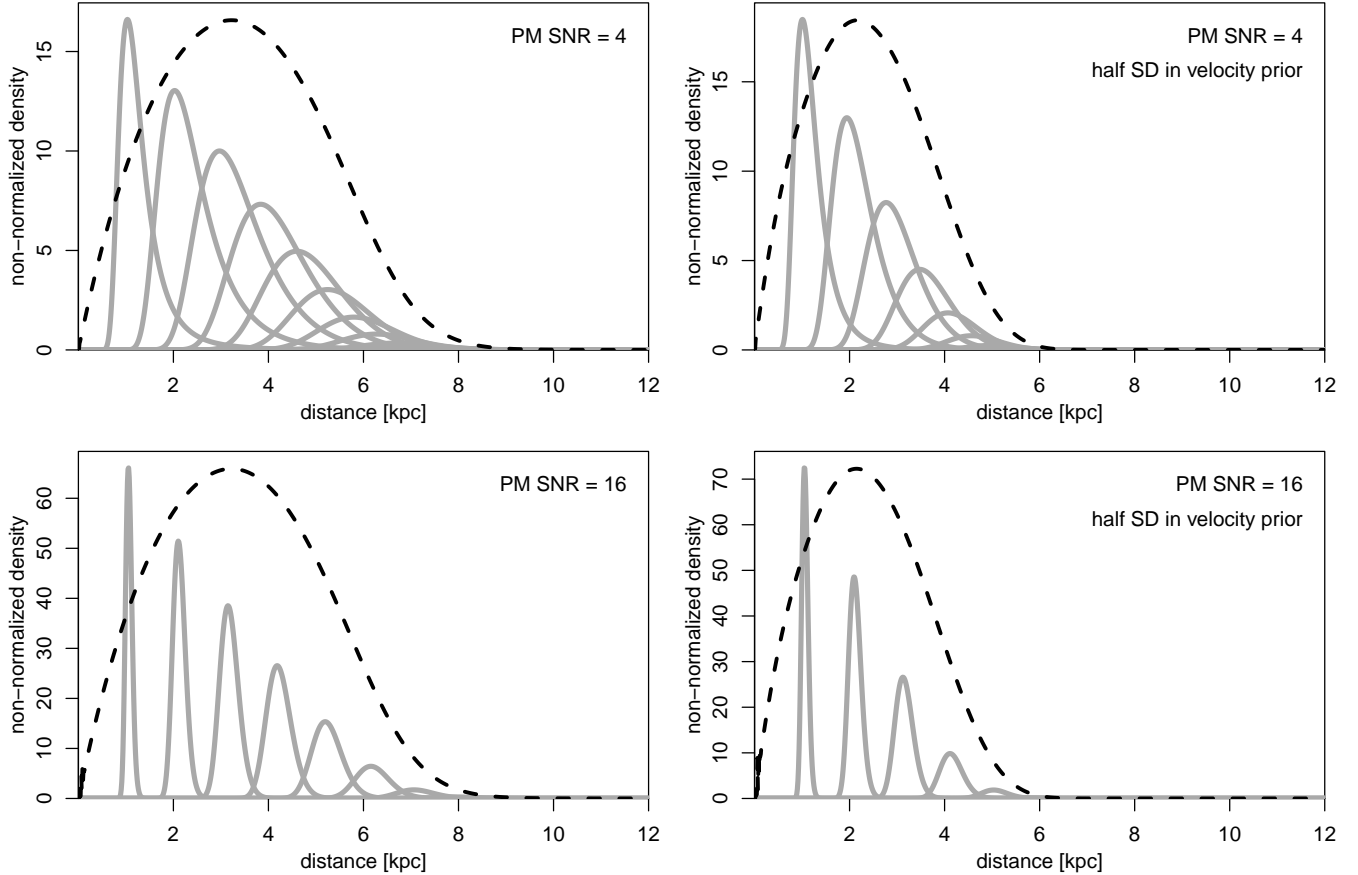
Figure 21. Demonstration of how the likelihood–prior product, $P(\mu \mid r, v)P(v \mid r)$, shown as the grey curves, sum to make the unnormalized distance posterior (black dashed line) via equation A1. The grey curves are shown just for 10 velocities from 5 to $50\,\mathrm{km\,s^{-1}}$ in steps of $5\,\mathrm{km\,s^{-1}}$, but some are so flat as to be invisible. In all panels the measured proper motion is $1\,\mathrm{mas\,yr^{-1}}$. The top row is for a proper motion SNR of 4, the bottom for 16. The the right column if for a narrower prior with half the standard deviation of that in the left column. None of the functions shown are probability density functions in r. The grey ones have been normalized to unit area over the range plotted. The black dashed curve has been scaled to have the same maximum as the largest grey peak.

top-left panel of Figure 21, where only a few of the grey curves with different velocities are shown, but the resulting integral (dashed black line) has been computed using a broad, dense grid of velocities. We see from this plot that even though each contribution to the integral is relatively narrow, their sum, and therefore the posterior, is broad. Consequently, even if we have higher SNR proper motions – and so narrower likelihoods and thus narrower grey distributions – the integral is more or less unchanged. This is demonstrated in the bottom left panel of Figure 21, where the SNR is increased by a factor of four. This is the main point: improving the proper motion SNR does not necessarily improve the distance accuracy. This is why we expect little improvement in kinegeometric distances from GDR3 to GDR5 (section 5). If the prior is narrower, then we can get a narrower posterior for given data. This is illustrated in the right columns of Figure 21, where the standard deviation of the velocity prior has been halved at all distances. The proper motion likelihoods are unchanged (compared to the panel on the left), but some are now more suppressed by the prior.

B. RANDOM ERRORS CAUSE BIAS

One might think that, because the priors are built from the mock catalogue, there should be no bias in the results on the mock catalogue in section 3 (for either distance estimate). This is incorrect for at least three reasons. First, the mock catalogue used in the performance assessment is noisy. Although the noise is symmetric (Gaussian) in the parallax, it is not symmetric in the distance due to their nonlinear relation ($r \sim 1/\varpi$; see Bailer-Jones 2015). Second,

the distance and velocity priors are fits, so do not represent all the variances in the mock data. We may in fact not want a prior tuned exactly to all the details and specific choices of the mock catalogue, because this would be a rather strong prior. Third, in reflecting the true Galaxy, the distance prior has the bulk of its probability mass within a few kpc. Stars beyond these distances tend to have low parallax SNRs, i.e. the data are relatively uninformative, so the prior will tend to pull the inferred distance below the true distance for these stars. This is why we see larger biases at larger distances in Figure 7. Nearby low parallax SNR stars will tend to have their distances pulled up by the prior for the same reason, but this is much less common because there are fewer nearby stars, and because nearby stars tend to have higher parallax SNRs on account of their larger parallaxes and brighter apparent magnitudes.

"Not using a prior" is not a solution to this problem, because a prior is invariably present implicitly. Indeed, if we just invert a parallax to estimate the distance, the biases are much larger (Bailer-Jones 2015). A prior uniform in distance or velocity is a poor choice, and much more biased than the present choice: A uniform distance prior is equivalent to assuming that the stellar number density drops as $1/r^2$ from the Sun, which is demonstrably wrong. A uniform velocity distribution relative to the Sun contradicts the existence of the different components of the Galaxy. Unless one is prepared to use highly tuned priors, low SNR data give low SNR inferences. The only other solution is to acquire better data.

REFERENCES

Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2022, A&A, 658, A91, doi: 10.1051/0004-6361/202142369

Andrae, R., Fouesneau, M., Sordo, R., et al. 2023, A&A, 674, A27, doi: 10.1051/0004-6361/202243462

Astraatmadja, T. L., & Bailer-Jones, C. A. L. 2016a, ApJ, 832, 137 (paper II), doi: 10.3847/0004-637X/832/2/137

—. 2016b, ApJ, 833, 119 (paper III), doi: 10.3847/1538-4357/833/1/119

Bailer-Jones, C. A. L. 2015, PASP, 127, 994 (paper I), doi: 10.1086/683116

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147 (paper V), doi: 10.3847/1538-3881/abd806

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, AJ, 156, 58 (paper IV), doi: 10.3847/1538-3881/aacb21

Binney, J., & Merrifield, M. 1998, Galactic Astronomy (Princeton University Press)

Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1, doi: 10.1051/0004-6361/202038192

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: 10.1086/670067

Fouesneau, M., Andrae, R., Dharmawardena, T., et al. 2022, A&A, 662, A125, doi: 10.1051/0004-6361/202141828

Gaia Collaboration. 2022, Gaia DR3, European Space Agency, doi: 10.5270/esa-qa4lep3

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1, doi: 10.1051/0004-6361/202243940

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759, doi: 10.1086/427976

Hemenway, M. K. 1975, AJ, 80, 199, doi: 10.1086/111732

Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021a, A&A, 649, A2, doi: 10.1051/0004-6361/202039709

Lindegren, L., Bastian, U., Biermann, M., et al. 2021b, A&A, 649, A4, doi: 10.1051/0004-6361/202039653

Luri, X., Brown, A. G. A., Sarro, L. M., et al. 2018, A&A, 616, A9, doi: 10.1051/0004-6361/201832964

Olivares, J., Sarro, L. M., Bouy, H., et al. 2020, A&A, 644, A7, doi: 10.1051/0004-6361/202037846

Rate, G., & Crowther, P. A. 2020, MNRAS, 493, 1512, doi: 10.1093/mnras/stz3614

Reid, M. J. 2022, AJ, 164, 133, doi: 10.3847/1538-3881/ac80bb

Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523, doi: 10.1051/0004-6361:20031117

Rybizki, J., Demleitner, M., Bailer-Jones, C. A. L., et al. 2020, PASP, 132, 074501, doi: 10.1088/1538-3873/ab8cb0

Sanders, J. L. 2023, MNRAS, 523, 2369, doi: 10.1093/mnras/stad1431

Schönrich, R., McMillan, P., & Eyer, L. 2019, MNRAS, 487, 3568, doi: 10.1093/mnras/stz1451

Strömberg, G. 1936, ApJ, 84, 555, doi: 10.1086/143783

Zari, E., Rix, H. W., Frankel, N., et al. 2021, A&A, 650, A112, doi: 10.1051/0004-6361/202039726

Zhang, X., Green, G. M., & Rix, H.-W. 2023, MNRAS, 524, 1855, doi: 10.1093/mnras/stad1941

Zonca, A., Singer, L., Lenz, D., et al. 2019, The Journal of Open Source Software, 4, 1298, doi: 10.21105/joss.01298