The Astrophysical Journal, 833:119 (6pp), 2016 December 10 doi:10.3847/1538-4357/833/1/119

ESTIMATING DISTANCES FROM PARALLAXES. III. DISTANCES OF TWO MILLION STARS IN THE *Gaia* DR1 CATALOGUE

Tri L. Astraatmadja[1,2] and Coryn A. L. Bailer-Jones[2]
[1] Department of Terrestrial Magnetism, Carnegie Institution for Science, 5241 Broad Branch Road, NW, Washington, DC 20015-1305, USA
[2] Max Planck Institute for Astronomy, Königstuhl 17, D-69117, Heidelberg, Germany
Received 2016 September 21; revised 2016 September 28; accepted 2016 September 29; published 2016 December 12

ABSTRACT

We infer distances and their asymmetric uncertainties for two million stars using the parallaxes published in the *Gaia* DR1 (GDR1) catalogue. We do this with two distance priors: A minimalist, isotropic prior assuming an exponentially decreasing space density with increasing distance, and an anisotropic prior derived from the observability of stars in a Milky Way model. We validate our results by comparing our distance estimates for 105 Cepheids which have more precise, independently estimated distances. For this sample we find that the Milky Way prior performs better (the rms of the scaled residuals is 0.40) than the exponentially decreasing space density prior (rms is 0.57), although for distances beyond 2 kpc the Milky Way prior performs worse, with a bias in the scaled residuals of −0.36 (versus −0.07 for the exponentially decreasing space density prior). We do not attempt to include the photometric data in GDR1 due to the lack of reliable color information. Our distance catalog is available at http://www.mpia.de/homes/calj/tgas_distances/main.html as well as at CDS. This should only be used to give individual distances. Combining data or testing models should be done with the original parallaxes, and attention paid to correlated and systematic uncertainties.

Key words: catalogs – methods: data analysis – methods: statistical – parallaxes – stars: distances – surveys

Supporting material: machine-readable tables

1. INTRODUCTION

The European Space Agency (ESA) *Gaia* mission (Gaia Collaboration et al. 2016b) is obtaining highly accurate parallaxes and proper motions of over one billion sources brighter than $G \simeq 20.7$. The first data release (*Gaia* DR1), based on early mission data, was released to the community on 2016 September 14 (Gaia Collaboration et al. 2016a). The primary astrometric data set in this release lists the positions, parallaxes, and proper motions of 2,057,050 stars, which are in the *Tycho*-2 (Høg et al. 2000) catalogue (93,635 of the these are *Hipparcos* (Perryman et al. 1997; van Leeuwen 2007) sources). This data set is called the *Tycho-Gaia* astrometric solution (TGAS; Michalik et al. 2015; Lindegren et al. 2016).

The five-parameter astrometric solutions for TGAS stars were obtained by combining *Gaia* observations with the positions and their uncertainties of the *Tycho*-2 stars (with an observation epoch of around J1991) as prior information (Lindegren et al. 2016). This was necessary, because the observation baseline in the early *Gaia* data was insufficient for a *Gaia*-only solution. The resulting solutions have median parallax uncertainties of ∼0.3 mas, with an additional systematic uncertainty of about ∼0.3 mas (Gaia Collaboration et al. 2016a; Lindegren et al. 2016).

Using the TGAS parallaxes ϖ and uncertainties σ_ϖ, we here infer the distances to all TGAS stars along with (asymmetric) distance uncertainties (as Bayesian credible intervals). The motivation and methods to estimate distances from parallaxes have been described in our earlier works (Bailer-Jones 2015; Astraatmadja & Bailer-Jones 2016, henceforth Papers I and II respectively). We will not repeat the discussion here, except to remind readers that inverting parallaxes to estimate distances is only appropriate in the absence of noise. As parallax measurements have uncertainties—and for many TGAS stars very large uncertainties—distance estimation should always be treated as an inference problem.

2. PROPERTIES OF TGAS PARALLAXES AND THEIR MEASUREMENT UNCERTAINTIES

Panels (a)–(e) of Figure 1 show the distribution of σ_ϖ as a function of ϖ, as well as histograms of ϖ and σ_ϖ. The distribution in σ_ϖ covers a narrow range between 0.2 and 1 mas (cf. Figure 13 of Paper II, which shows the same plot for GUMS data[3]), which reflects the preliminary nature of GDR1. The upper limit of 1 mas is due to the imposed $\sigma_\varpi = 1$ mas cutoff to reject unreliable astrometric solutions, while the lower limit is due to the ∼0.2 mas noise floor, which is dominated by the satellite attitude and calibration uncertainties (Lindegren et al. 2016). Future data releases will much more precise (Gaia Collaboration et al. 2016a).

In Panels (f)–(g) of Figure 1, we show the distribution of the fractional parallax uncertainties $f_{\rm obs} = \sigma_\varpi/\varpi$ of TGAS stars, compared with *Hipparcos* and GUMS stars. It can be seen here that, interestingly, the combination of TGAS ϖ and σ_ϖ produces a distribution of $f_{\rm obs}$, which is similar to that of *Hipparcos* stars.

3. METHOD, PRIORS, AND DATA PRODUCTS

The inferred distances of stars depend, not only on the observed parallaxes and their uncertainties, but on the prior. In this paper, we infer distances using two priors: a minimalist, isotropic exponentially decreasing space density prior and a more complex, anisotropic Milky Way prior. The properties of the exponentially decreasing space density have been discussed in Paper I, and in Paper II we have seen that for an end-of-

[3] GUMS, the *Gaia* Universe Model Snapshot (Robin et al. 2012), is a mock catalog, which simulates the expected content of the final *Gaia* catalogue.

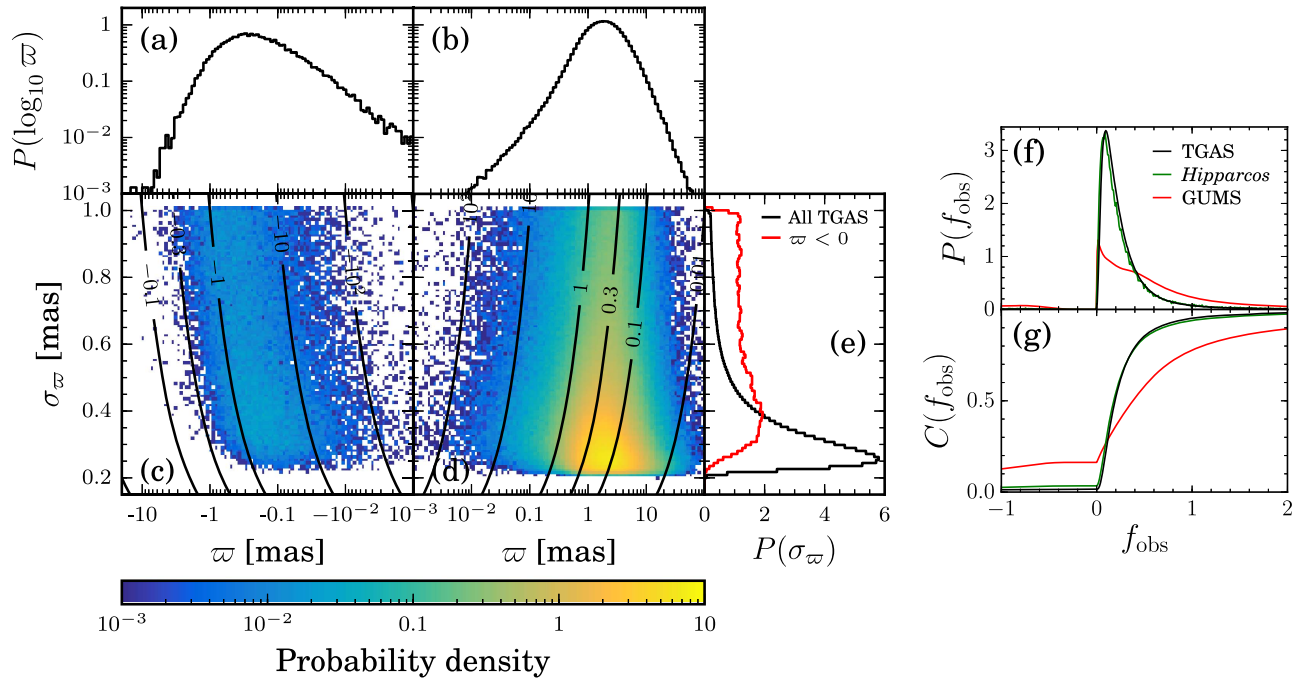

Figure 1. TGAS parallax data. Panels (a) and (b) show the histograms of the TGAS parallaxes, ϖ, for negative and positive parallaxes, respectively. Panels (c) and (d) show the distribution of TGAS parallax uncertainties, σ_ϖ, as a function of ϖ, on a density scale, again for negative and positive parallaxes. The contour lines show the loci of constant $f_{\rm obs} = \sigma_\varpi/\varpi$ as indicated by the labels. Panel (e) shows the histogram of σ_ϖ for all stars (black histogram) as well as the subset, which have negative parallaxes (red histogram). The histogram for only positive parallaxes is almost exactly the same as those for all stars and thus is not shown. The vertical axis in Panels (a) and (b) is logarithmic, whereas it is linear in Panels (c)–(e). Panel (f) shows the probability density of the observed fractional parallax uncertainty, $f_{\rm obs} = \sigma_\varpi/\varpi$, for TGAS stars (black line), compared with *Hipparcos* (green line) and GUMS (red line) stars. Panel (g) shows the corresponding cumulative distributions. Note that Panels (f)–(g) only cover a subrange of all possible $f_{\rm obs}$.

mission *Gaia*-like catalog, the optimum scale length L is 1.35 kpc. We use this value to derive distances here, even though it is optimized for the end-of-mission catalog, so TGAS stars may have a different true distance distribution.

Although not analyzed here, in our catalog we also provide distances using the exponentially decreasing space density prior using $L = 0.11$ kpc. This value is found by fitting the prior with the true distance distribution of GUMS stars with $V < 11$ (this is the V-band magnitude at which *Tycho*-2 is 99% complete).

The derivation and parameters of the Milky Way prior have been discussed in Paper II, and illustrations of the resulting posterior for several parallaxes ϖ and uncertainties σ_ϖ can be seen in Figures 6 and 7 of Paper II. Here, we retain the parameters of the Milky Way model as well as the Drimmel et al. (2003) extinction map, with the exception of the limiting magnitude $m_{G,\rm lim}$ (Equation (6) in Paper II), used to calculate the faint end of the luminosity function. In this paper, we use $m_{G,\rm lim} = 12.998$, which is the 99.9% percentile of the magnitude distribution of all TGAS stars.

For every single star, we compute the posterior probability density function over distance. The distance estimate we report here is the mode of the posterior, $r_{\rm Mo}$. We do not report the median distance, because as we have seen in Paper II, it is a worse estimator for the priors used here.

In addition to the median, we report in our catalog the 5% and 95% quantiles of the posterior, r_5 and r_{95}. Note that many of the posteriors are asymmetric about the mode (and mean and median). The difference between these gives a 90% credible interval, which we then divide by a factor $2s$ to produce

$$\sigma_r = \frac{r_{95} - r_5}{2s}, \quad (1)$$

where $s = 1.645$ is ratio of the 90%–68.3% credible interval *in a Gaussian* distribution. Thus, σ_r is a simplified (symmetric) uncertainty in our distance estimate, which is equivalent, in some sense, to a 1σ Gaussian uncertainty.

We use neither apparent magnitudes nor colors to help infer the distance, even though we have shown in Paper II that this significantly improves the distance estimation in many cases. This is because GDR1 does not contain color information. We chose not to use the *Tycho* photometric data on account of its low precision (median photometric uncertainties in B_T and V_T are, respectively, 136 and 96 mmag).

In the analyses that follow, we have not included in our inference the ~0.3 mas systematic uncertainties reported for the TGAS parallaxes. This is partly because we know this to be a very rough estimate of the systematics, and is possibly overestimated. But we do provide a second catalog on the website mentioned, which includes this systematic error. It is included by adding it in quadrature with the random parallax error and then repeating the inference. In general, this affects both the mode of the posterior (the distance estimate) and its quantiles (the uncertainty).

4. DISTANCE ESTIMATION RESULTS

The results of the distance estimation are shown in Figure 2 and the statistics of the uncertainties are summarized in Table 1.

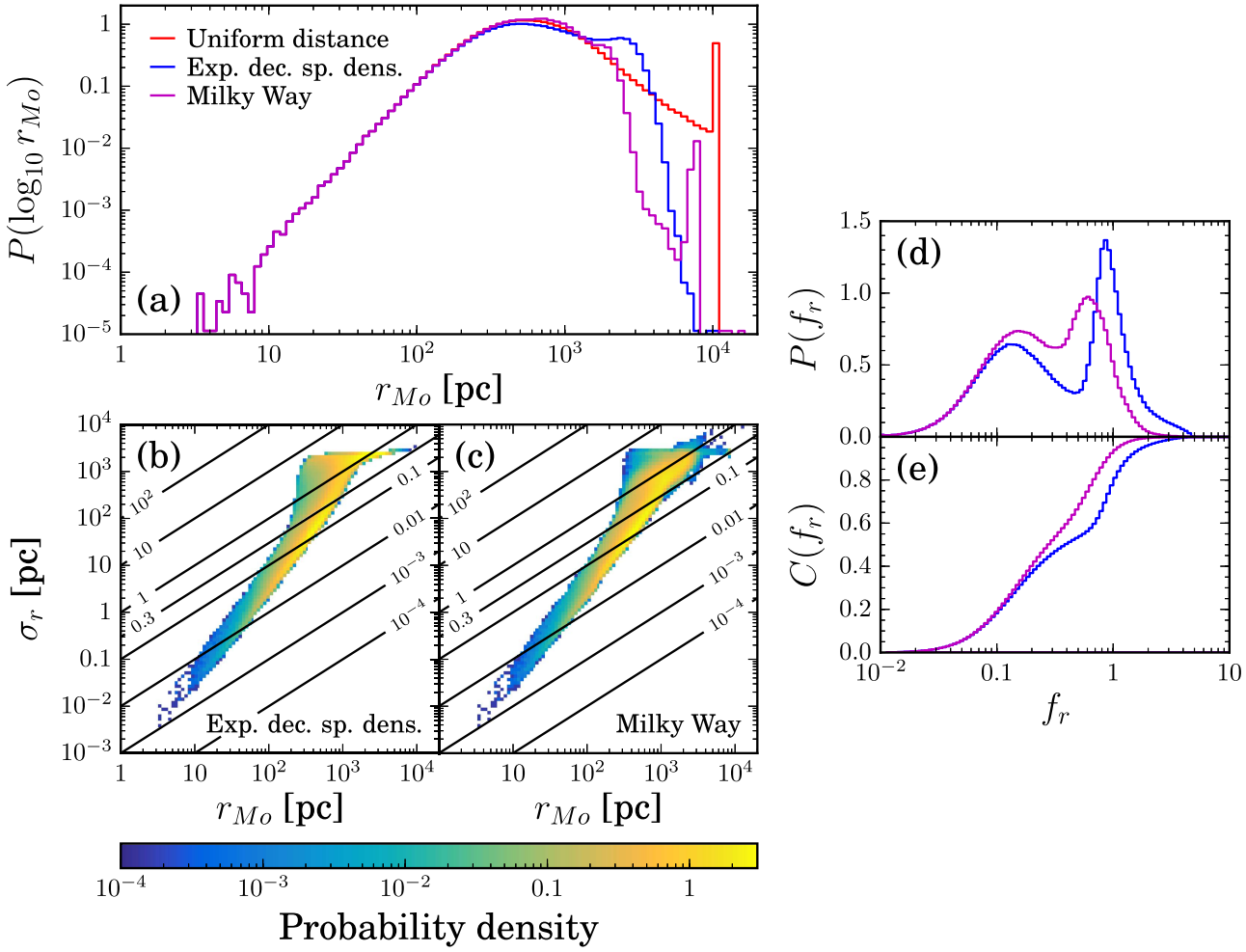

Figure 2. Distance estimates for TGAS stars. Panel (a) shows the distribution of the estimated distances $r_{\rm Mo}$ derived from the mode of the three posteriors indicated by the legend. Panels (b) and (c) show the distribution of the distance uncertainties, σ_r, as a function of $r_{\rm Mo}$, on a density scale, for the exponentially decreasing space density and the Milky Way priors, respectively. The diagonal lines show the loci of constant $f_r = \sigma_r/r_{\rm Mo}$ as indicated by the labels. Panel (d) shows the probability distribution functions of the fractional distance uncertainty $f_r = \sigma_r/r_{\rm Mo}$ for the exponentially decreasing space density prior (blue) and the Milky Way prior (magenta). Panel (e) shows the corresponding cumulative distributions.

Table 1
Statistical Summary of the Distance Estimation of Two Million Sources in the Primary Data Set of GDR1

Data set	TGAS			*Hipparcos* Subset		
	10%	50%	90%	10%	50%	90%
Exponentially Decreasing Space Density						
All stars	0.067	0.378	1.315	0.021	0.078	0.656
$r_{\rm Mo} < 200$ pc	0.023	0.045	0.095	0.021	0.077	0.365
Milky Way						
All stars	0.066	0.273	0.874	0.021	0.077	0.365
$r_{\rm Mo} < 200$ pc	0.023	0.046	0.096	0.013	0.035	0.069

Note. Columns with headings 10%, 50%, and 90% give the lower decile, median, and upper decile of the fractional uncertainty f_r for all 2,057,050 sources in the primary data set as well as a subset of 93,635 sources in common with *Hipparcos*.

In Panel (a) of Figure 2, we show the distribution of the estimated distance $r_{\rm Mo}$ derived from the mode of the two posteriors already mentioned. The red line in that panel is for a third posterior, which uses the uniform distance prior (Paper I), with a large cutoff at $r_{\rm lim} = 10$ kpc. This posterior is equivalent to inverting the parallax to get a distance, except for the cases where the parallax is very small or negative, in which case the mode of the posterior is at $r_{\rm lim} = 10$ kpc. This is the reason for the peak in the distribution we see in Panel (a). It contains 43,673 stars, which is 2.1% of TGAS. For the exponentially decreasing space density prior, we also see a peak, but at around $r_{\rm Mo} = 2.7$ kpc (it is not very visible as a peak due to the log scale). This is the mode of that prior ($r = 2L$), and the mode of the posterior is very close to this for stars with large parallax uncertainties. The Milky Way prior also has a mode, but because it is an anisotropic prior, the mode varies with line-of-sight direction. However, the most prominent peak at $r_{\rm Mo} \sim 8$ kpc can be seen, which corresponds to the prior for stars toward the Galactic center, and thus for poorly measured stars in this direction.

For distances up to about 200 pc, the distributions of $r_{\rm Mo}$ for both priors are similar to each other. Looking again at Panel (d) of Figure 1, we see that for stars with $\varpi \gtrsim 5$ mas, most have $f_{\rm obs} < 0.2$. We showed in Paper II that for stars with positive parallaxes and $f_{\rm obs} \lesssim 0.2$, the distance estimate is largely

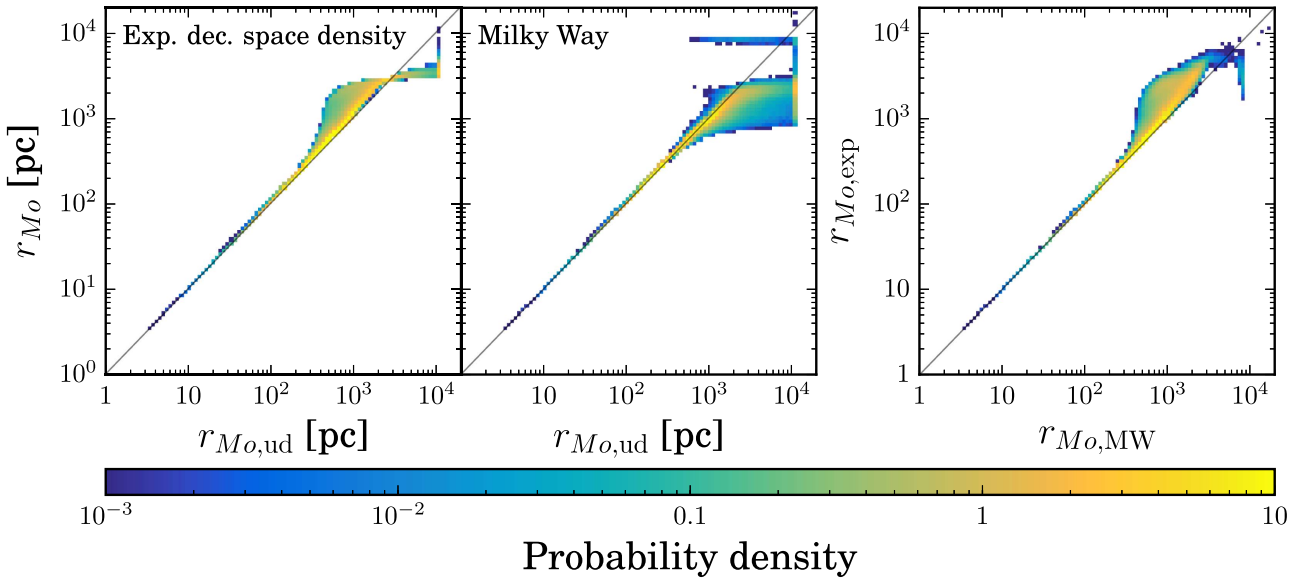

Figure 3. Distance estimate comparisons. The left two panels compare our distance estimates from the exponentially decreasing space density and Milky Way priors (vertical axes) with those obtained from the naive uniform distance prior (horizontal axis). The right panel compares the estimates from our two pain priors, the exponentially decreasing space density prior (vertical axis) and Milky Way prior (horizontal axis).

independent of the choice of prior. Beyond 200 pc, however, the r_{Mo} distributions for all priors diverge. For distances of more than 1 kpc, most stars have $f_{\mathrm{obs}} \gtrsim 0.3$ and the distance estimate becomes much more prior-dependent.

The distribution of the fractional uncertainties in distance $f_r = \sigma_r / r_{\mathrm{Mo}}$ is shown in Panels (d) and (e) of Figure 2. For both priors, the combined distribution of r_{Mo} and σ_r is similar for $f_r \lesssim 0.1$. Both distributions peak at about 0.15, but beyond that a second peak corresponding to poorly measured stars can be seen at $f_r \sim 0.8$ and $f_r \sim 0.6$ for the exponentially decreasing space density prior and the Milky Way prior, respectively.

We compare the distances estimated using the two priors with each other, and with distances estimated from the uniform distance prior, in Figure 3. We see again that for distances up to ~200 pc, distances using all priors are similar. For $1/\varpi \gtrsim 200$ pc, we start to see elongations that correspond to the mode of the respective priors, as discussed above.

5. VALIDATION WITH CEPHEID VARIABLES

To see how consistent our estimated distances are with other, more precise, estimates (for distant stars), we compare our estimated distances with the distances of Cepheid variable stars. We took 170 Cepheids from Groenewegen (2013) and cross-matched them with GDR1, using Simbad. We found 105 Cepheids in common with GDR1. The Groenewegen (2013) Cepheids have median fractional uncertainties of about ~0.054. Almost all of these Cepheids are *Hipparcos* sources.

Figure 4 compares our distance estimates (for both priors) with those of Groenewegen (2013) for both priors. The bottom row of that figure shows this using the scaled differences

$$x_{\mathrm{Mo}} = \frac{r_{\mathrm{Mo}} - r_{\mathrm{Cep}}}{r_{\mathrm{Cep}}}. \quad (2)$$

The uncertainties in r_{Cep} are taken from Groenewegen (2013), where they were computed in a Monte Carlo simulation, which takes into account uncertainties in the spectrophotometry, the projection factor, and the phase measurements. We multiply these uncertainties by $s = 1.645$ to scale them to be 90% credible intervals, in order to make a fair comparison with our 90% credible intervals, $r_{95} - r_5$.

To summarize the differences seen in Figure 4, we calculate the bias $\bar{x}$, rms $\overline{x^2}^{1/2}$ of the scaled residuals, as well as the standard deviation $\sigma_{x,\mathrm{Mo}}$ of the scaled residuals, for all Cepheids, for both priors. We also do this separately for near ($r_{\mathrm{Cep}} < 2$ kpc) and distant ($r_{\mathrm{Cep}} \geqslant 2$ kpc) Cepheids. These results are summarized in Table 2.

Inspecting Figure 4 and Table 2, and assuming the Groenewegen (2013) distances to be "true" (for simplicity), we see that overall the Milky Way prior performs better than the exponentially decreasing space density prior in terms of having a smaller rms and standard deviation. It is slightly less biased than the exponentially decreasing space density prior, although the bias is in the opposite direction: it tends to underestimate distance. This is due to the assumptions the Milky Way prior makes in the face of poor data, which is that a star is more likely to reside in the disk than further away. Hence, this prior becomes mismatched when we only consider the distant Cepheids ($r_{\mathrm{Cep}} \geqslant 2$ kpc). Distance estimates using the Milky Way prior have a bias of -0.36 for these stars, as is also apparent from Figure 4. For $r_{\mathrm{Cep}} \geqslant 2$ kpc, when the data are poor, the posterior based on this prior has a mode at around about 2 kpc, which roughly corresponds to the radial scale length of the thick disk in our Milky Way model. For Cepheids closer than 2 kpc, however, we see that the Milky Way prior performs well in terms of bias, rms, and standard deviation.

The Milky Way prior also gives a more reasonable credible interval than the exponentially decreasing space density prior, as can be seen in the top row of Figure 4. Most of our TGAS-based distance uncertainties are large, because the Cepheids are distant and have large fractional parallax uncertainties, with median f_{obs} of about 0.48 (versus ~0.2 for all TGAS stars). Furthermore, the posteriors—and, therefore, the credible intervals—are highly asymmetric, with a long tail to large

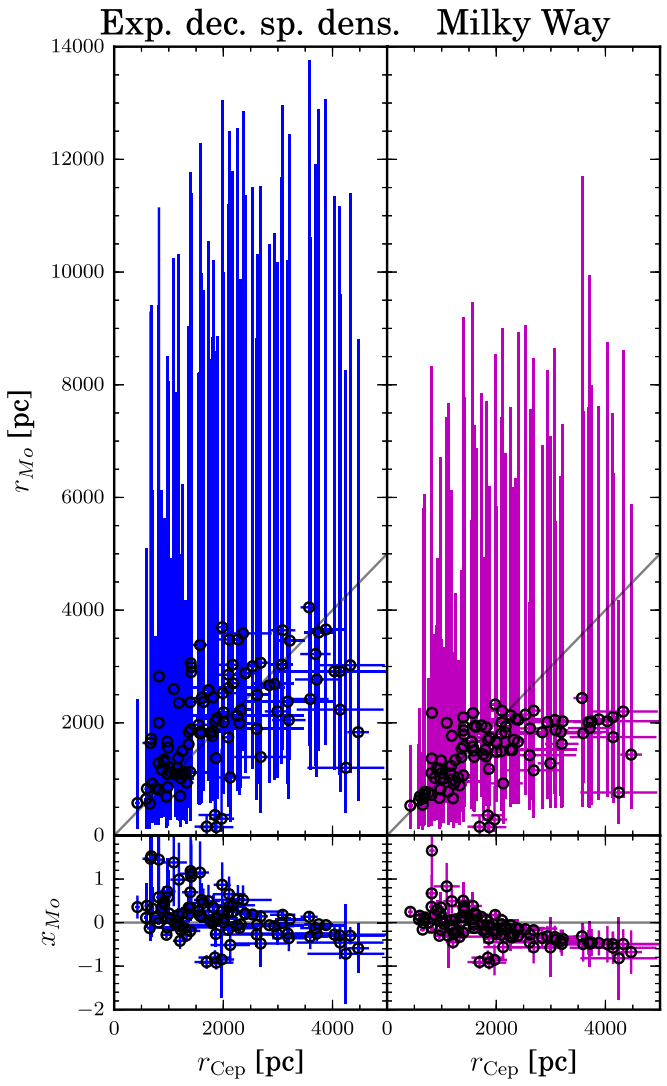

Figure 4. A comparison of the distances estimated using the period–luminosity relation of Cepheid stars (Groenewegen 2013) in common with TGAS sources, with the distances of the same stars estimated using the exponentially decreasing space density (left column) and the Milky Way (right column) prior. The top row shows the distance comparisons. The diagonal lines indicate a perfect match between the distances. The bottom row shows the scaled residual x_{Mo} as a function of the Cepheid distance. The horizontal lines indicate zero residuals. The error bars of the estimated distances are the 90% credible intervals, while for the Cepheids they are the quoted 1σ uncertainties multiplied by $s = 1.645$ to scale them into the 90% credible intervals.

Table 2
The Bias $\overline{x}$ as Well as the rms $\overline{x^2}^{1/2}$ and Standard Deviation $\sigma_{x,\rm Mo}$ of the Scaled Residuals of Cepheids Stars in the TGAS Catalogue

Prior and sample	$\overline{x}$	$\overline{x^2}^{1/2}$	$\sigma_{x,\rm Mo}$
Exponentially decreasing space density			
All Cepheids	0.151	0.567	0.547
Cepheids with $r_{\rm Cep} < 2$ kpc	0.298	0.678	0.608
Cepheids with $r_{\rm Cep} \geqslant 2$ kpc	−0.070	0.340	0.333
Milky Way			
All Cepheids	−0.133	0.404	0.382
Cepheids with $r_{\rm Cep} < 2$ kpc	0.022	0.395	0.394
Cepheids with $r_{\rm Cep} \geqslant 2$ kpc	−0.364	0.418	0.205

Table 3
Distances and Uncertainties of All TGAS Sources

Column	Description
1	*Hipparcos* identifier
2	*Tycho*-2 identifier
3	Source identifier
4	Galactic longitude at epoch 2015.0
5	Galactic latitude at epoch 2015.0
6	Absolute barycentric stellar parallax at the reference epoch
7	Standard error absolute barycentric stellar parallax
8	*Gaia G*-band mean magnitude
9	Mode distance of posterior using the exponentially decreasing space density prior with $L = 0.11$ kpc
10	5th percentile of posterior using the exponentially decreasing space density prior with $L = 0.11$ kpc
11	50th percentile of posterior using the exponentially decreasing space density prior with $L = 0.11$ kpc
12	95th percentile of posterior using the exponentially decreasing space density prior with $L = 0.11$ kpc
13	Standard error in distance using the exponentially decreasing space density prior with $L = 0.11$ kpc
14	Mode distance of posterior using the exponentially decreasing space density prior with $L = 1.35$ kpc
15	5th percentile of posterior using the exponentially decreasing space density prior with $L = 1.35$ kpc
16	50th percentile of posterior using the exponentially decreasing space density prior with $L = 1.35$ kpc
17	95th percentile of posterior using the exponentially decreasing space density prior with $L = 1.35$ kpc
18	Standard error in distance using the exponentially decreasing space density prior with $L = 1.35$ kpc
19	Mode distance of posterior using Milky Way prior
20	5th percentile of posterior using Milky Way prior
21	50th percentile of posterior using Milky Way prior
22	95th percentile of posterior using Milky Way prior
23	Standard error in distance using Milky Way prior

(This table is available in its entirety in machine-readable form.)

distances. This is a natural consequence of the nonlinear transformation from parallax to distance.

The stars used in this validation are intrinsically bright and relatively distant compared to the typical Milky Way stars used to build the Milky Way prior. Our distance estimation is based solely on measured parallaxes; no photometry is involved. Thus, the Milky Way prior is not well-matched: in the absence of precise parallaxes it tells us that stars are more likely to be in the disk than further away. This explains the poorer behavior of this prior for distant Cepheids. The exponentially decreasing space density performs better in this regime due the scale length L adopted, which puts the mode of the prior at $2L = 2.7$ kpc.

6. CONCLUSIONS

We have inferred the distances of two million stars in the *Gaia* DR1 catalogue using Bayesian inference. The priors used are the exponentially decreasing space density prior with scale length $L = 1.35$ kpc, and the Milky Way prior with the same parameters as in Paper II. The median fractional distance uncertainties ($f_r = \sigma/r_{\rm Mo}$) are 0.38 and 0.27 for the exponentially decreasing space density and the Milky Way prior, respectively. If we only consider stars with the estimated

Table 4
Distances and Uncertainties of all TGAS Sources

Column	Description
1	*Hipparcos* identifier
2	*Tycho*-2 identifier
3	Source identifier
4	Galactic longitude at epoch 2015.0[a]
5	Galactic latitude at epoch 2015.0[a]
6	Absolute barycentric stellar parallax at the reference epoch[b]
7	Standard error absolute barycentric stellar parallax
8	*Gaia G*-band mean magnitude
9	Mode distance of posterior using the exponentially decreasing space density prior with $L = 0.11$ kpc[c]
10	5th percentile of posterior using the exponentially decreasing space density prior with $L = 0.11$ kpc[c]
11	50th percentile of posterior using the exponentially decreasing space density prior with $L = 0.11$ kpc[c]
12	95th percentile of posterior using the exponentially decreasing space density prior with $L = 0.11$ kpc[c]
13	Standard error in distance using the exponentially decreasing space density prior with $L = 0.11$ kpc[c]
14	Mode distance of posterior using the exponentially decreasing space density prior with $L = 1.35$ kpc[d]
15	5th percentile of posterior using the exponentially decreasing space density prior with $L = 1.35$ kpc[d]
16	50th percentile of posterior using the exponentially decreasing space density prior with $L = 1.35$ kpc[d]
17	95th percentile of posterior using the exponentially decreasing space density prior with $L = 1.35$ kpc[d]
18	Standard error in distance using the exponentially decreasing space density prior with $L = 1.35$ kpc[d]
19	Mode distance of posterior using Milky Way prior[e]
20	5th percentile of posterior using Milky Way prior[e]
21	50th percentile of posterior using Milky Way prior[e]
22	95th percentile of posterior using Milky Way prior[e]
23	Standard error in distance using Milky Way prior[e]
24	Groenewenegen 2013 *Hipparcos* identifier
25	Groenewenegen 2013 *Tycho*-2 identifier
26	Groenewenegen 2013 Cepheid distance
27	Error in the fit of Groenewenegen 2013 Cepheid distance
28	Error based on Monte Carlo simulation of Groenewenegen 2013 Cepheid distance

Notes.
[a] Aat epoch 2015.0.
[b] At the reference epoch.
[c] Using the exponentially decreasing space density prior with $L = 0.11$ kpc.
[d] Using the exponentially decreasing space density prior with $L = 1.35$ kpc.
[e] Using Milky Way prior.

(This table is available in its entirety in machine-readable form.)

distances of $r_{\rm Mo} < 200$ pc, the median value of f_r improves to about ∼0.04 for both priors. This applies to about 193,000 stars (the exact number is different for both priors) or about 9% of TGAS.

We validate our distance estimates using more precise distances for Cepheid stars in TGAS taken from Groenewegen (2013). We found that for distances closer than 2000 pc, the Milky Way prior performs better than the exponentially decreasing space density prior. Beyond 2000 pc, the Milky Way prior performs worse for this sample (which are intrinsically bright and distant stars), because it assumes that stars are more likely to be closer in the disk than further away. Our exponentially decreasing space density prior has a longer scale length and thus performs better in this sample, when faced with the same poor measurements. But overall, the Milky Way prior performs better.

Due to the lack of reliable colors, we do not use these in combination with the parallaxes to estimate distances. Rather than using the *Tycho* magnitudes, significant improvements can be achieved taking spectrophotometric information from other surveys. We choose here just to present astrometric distances.

The distance estimates presented in this paper are useful for *individual* stars. To obtain the mean distance to a group of stars, such as a cluster, one should do a combined inference using the original parallaxes and taking into account the correlated parallax uncertainties for stars observed in a small field. Note, however, that this combination will still not reduce the uncertainty in the mean below the limit presented by the TGAS systematic parallax error. Similarly, if one wishes to compare a model for distances to the TGAS data, this is normally best done by projecting the model-predicted distances into the parallax domain, rather than using individual estimated distances.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (http://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, http://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. We also made use of NASA's Astrophysics Data System; the SIMBAD database, operated at CDS, Strasbourg, France.

Software: `matplotlib` (Hunter 2007), `TOPCAT` (Taylor 2005).

APPENDIX
DATA CATALOGS

The appendix contains the distances and uncertainties of all TGAS sources in Table 3 and for a subset, the Groenewegen (2013) Cepheids 4. These distance catalogs are also available at http://www.mpia.de/homes/calj/tgas_distances/main.html.

REFERENCES

Astraatmadja, T. L., & Bailer-Jones, C. A. L. 2016, ApJ, 832, 137
Bailer-Jones, C. A. L. 2015, PASP, 127, 994
Drimmel, R., Cabrera-Lavers, A., & López-Corredoira, M. 2003, A&A, 409, 205
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016a, A&A, 595, 2
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016b, A&A, 595, 1
Groenewegen, M. A. T. 2013, A&A, 550, A70
Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27
Hunter, J. D. 2007, CSE, 9, 90
Lindegren, L., Lammers, U., Bastian, U., et al. 2016, A&A, 595, 4
Michalik, D., Lindegren, L., & Hobbs, D. 2015, A&A, 574, A115
Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49
Robin, A. C., Luri, X., Reylé, C., et al. 2012, A&A, 543, A100
Taylor, M. B. 2005, in ASP Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
van Leeuwen, F. 2007, Hipparcos: the New Reduction of the Raw Data, Vol. 350 Astrophysics and Space Science Library