Improving distance estimates for GDR3

prepared by: C.A.L. Bailer-Jones
reference: GAIA-C8-TN-MPIA-CBJ-089
issue: 1
revision: 0
date: 2020-09-29
status: Issued

Abstract

I consider how to provide improved distance estimates for GDR3, taking into account experience on the use of the distances provided for GDR2. I consider in particular the provision of transform-invariant distance estimates, distance moduli, and the use of alternative priors, such as the exponentially-decreasing distance (EDD) prior and a prior on absolute magnitude.

Document History

Issue	Revision	Date	Author	Comment
D	1	2019-04-24	CBJ	First draft
D	2	2019-11-19	CBJ	Comments from RAN, JR, MF. Added bullet items and figures to section 4; added section 5 and figures therein; added Figure 23; small revisions of text; updated the conclusions; added the recommendations (section 11)
D	3	2019-11-25	CBJ	Incorporated suggestions from the MPIA group
D	4	2020-01-04	CBJ	Added sections 6 and 8
1	0	2020-09-29	CBJ	Released without further modification

Contents

1 Scope

I consider what distance estimates we may want to provide with – or infer from – GDR3 (taken here to mean the early or full release). I focus on geometric distances for all sources with parallaxes, because spectrophotogeometric distances will anyway be estimated within GSP-Phot for stars with BP/RP spectra, G-band fluxes, and parallaxes (see, for example, CBJ-087).

For easy reference I frequently refer to the fractional parallax uncertainty (fpu), which is defined as σ_ϖ/ϖ. This is negative when the parallax is negative, but when I refer to a "small" fpu I mean positive and small in magnitude For this reason the inverse of the fpu, the parallax signal-to-noise ratio, is sometimes more useful (this too can be negative).

2 The story so far

For the TGAS subset of GDR1, Astraatmadja & Bailer-Jones (2016b; hereafter paper III) estimated distances for two million stars using two priors: (i) the exponentially-decreasing space density (EDSD) prior with a fixed length scale L=1.35 kpc (and additionally using L=0.11 kpc), and (ii) a Milky Way prior built from a 3D multi-component Milky Way model that included a dust model and a TGAS selection function, conditioned on l, b (described in detail in paper II). In the first case we just used a (Gaussian) likelihood in parallax. In the second case we included also a likelihood on the apparent magnitude and colours, which necessitated the adoption of a prior on the absolute magnitude and extinction (in order to link the magnitude to the distance), i.e. it involved assumptions about the intrinsic properties of stars. In both cases we did not apply a parallax zero point offset and the parallax uncertainties were used as provided.[1] We reported both the mode and median of the posterior PDFs as point estimates, along with 5% and 95% equal-tailed intervals as uncertainty estimates. There were, therefore, six distance estimates (with three corresponding confidence intervals) per star.

For GDR2, Bailer-Jones et al. (2018; hereafter paper IV) estimated distances for 1.33 billion stars using the EDSD prior with a sky-position dependence, $L(l, b)$, built using a (different) Galaxy model which incorporated both a dust model and a GDR2 selection function (the GDR2 mock catalogue, described in Rybizki et al. 2018). The likelihood was again Gaussian in parallax. There was no dependence on colours or magnitudes. A global parallax zeropoint of -0.029 mas was used and parallax uncertainties were not inflated. We summarized the posterior using the mode and the 68% highest density confidence interval (HDI). For 1.2 million sources the posterior was bimodal, and for 0.9 million sources the HDI could not be found (these two sets have significant overlap; see Table 3 of paper IV). For this latter set we reported the median and equal-tailed 68% confidence interval (ETI; also called the central confidence interval) instead of the mode and HDI. For 3278 sources we could not compute either the HDI or the ETI (due to numerical issues with the posterior), so they were omitted. The final catalogue therefore had distance (and uncertainty) estimates to every source in GDR2 that had a parallax, except

[1] We actually provided a second table which included a systematic error of 0.3 mas added in quadrature, but this is incorrect – the systematic is already accommodated in the uncertainties – so should not be used.

these 3278 sources.

3 Reflections on what was done for GDR2

The paper IV distances have been widely used and the paper widely cited. Based both on a non-exhaustive reading of the citing papers, as well as some reflection, a few points can be made.

1. The paper IV prior, which is characterized by a single length scale L, has a mode that varies between between 768 pc and 4778 pc (1st and 99th percentiles of the distribution) with a median of 2947 pc (the prior mode is $2L$). For comparison, the estimated distances vary from 319 pc to 7317 pc (1st and 99th percentiles) with a median of 2841 pc. As only 11.8% of GDR2 parallaxes have a SNR greater than 5, many of the inferred distances are dominated by the prior. (This is of course unavoidable as long as we only use parallaxes to infer distances.) One consequence of this became noticeable in the publications. If a star has a *true* distance which is large (more than, say, 5 kpc), then its fpu is usually large (e.g. 0.1 mas $\times$ 5 kpc = 0.5). The distance estimate will therefore be dominated by the prior, and this has a mode which is smaller than the true distance. As a result we will tend to underestimate distances to stars which are truly distant (and we will only know this for the few cases where we have an independent distance estimate, e.g. spectrophotometric distances for red clump stars). This effect increases rapidly with distance, as both the parallax drops and the star gets fainter (so σ_ϖ also gets larger). This issue in principle also affects stars which have true distances that are much smaller than the prior mode, where we would tend to underestimate distances. Yet this is much less apparent, because such nearby stars usually have small fpu and so the prior plays little role. An exception is intrinsically very faint stars, such as ultra cool dwarfs, that will have a larger fpu despite their proximity.

2. The mode of a distribution is not invariant under (monotonic) transformations of the distribution. Consequently, the mode of the distance posterior is not at the same distance as the mode of the corresponding log distance posterior (where "corresponding" means "using the same prior over distance"). This is easy to see from

$$P(\ln r) = rP(r) \tag{1}$$

$$\frac{dP(\ln r)}{dr} = P(r) + r\frac{dP(r)}{dr} \,. \tag{2}$$

   An inconsistency therefore arises if we take the published distances and convert them into a distance modulus: this would not correspond to the maximum of the posterior distribution in distance modulus (for the same prior). The published HDI would also no longer be the HDI for the log distance (although it would still be *a* 68% confidence interval). The mode in $\log(r)$ can be computed by solving a cubic equation. It turns out to be very similar to the one for the mode in r (just replace $-2r^2$ in equation 19 of paper I with $-3r^2$). Yet the two modes can be very different, as illustrated in Figure 1.

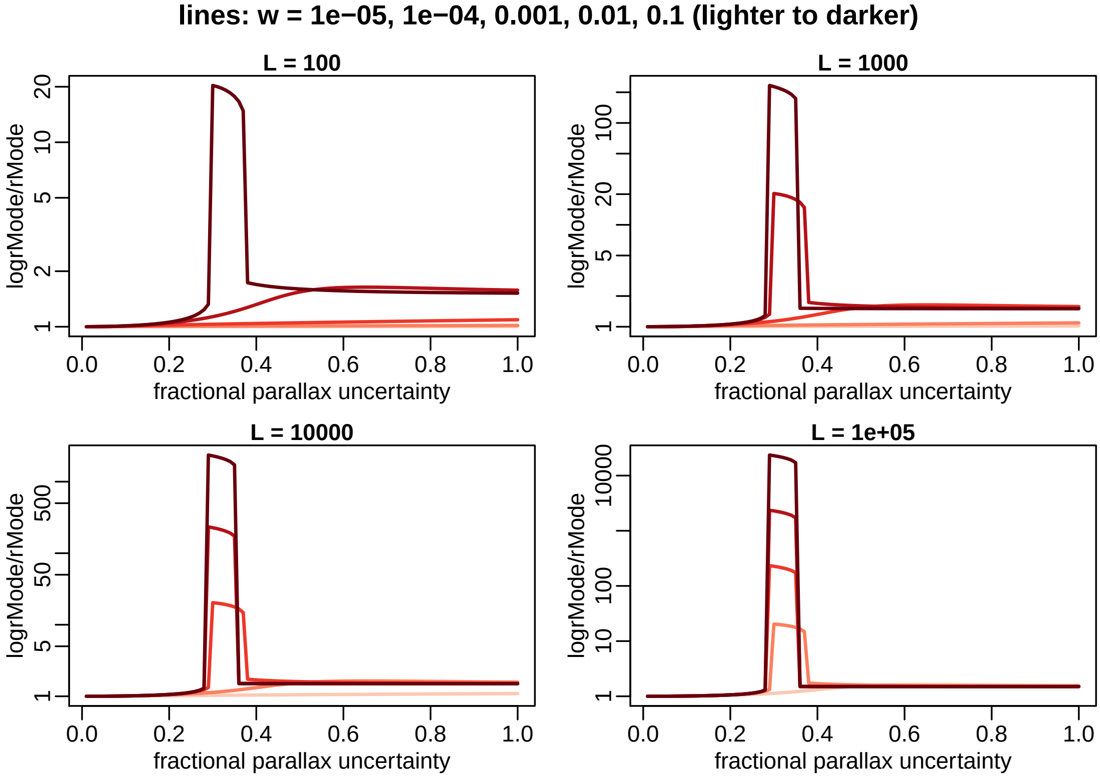

Figure 1: The ratio of [the mode of the posterior in log distance] to [the mode of the posterior in distance] for the EDSD prior, as a function of fractional parallax uncertainty (y-axis), for different parallaxes (in arcseconds, line colours) and length scales (panels).

3. Given that quite a few applications actually want to infer absolute magnitudes (via the distance modulus), sensible estimates of the log distance should be provided. Note that unlike the mode (and the mean), quantiles (such as the median) do transform monotonically. This will be discussed and explore below.

4. Some works that cite paper IV do not use its distances with the justification that the fpu is small (0.2 often being taken as the limit), and that therefore a parallax inversion is adequate. This is somewhat ironic given that, in the limit of small fpu, the distance estimate tends towards the inversion of the parallax anyway. Perhaps those authors felt that some "residual bias" from adopting the $L(l,b)$ prior was worse than the bias from (implicitly) adopting a uniform distance prior, even though the opposite is arguably the case.

4 Improving the distance estimates

The choice of prior always involves a trade-off, and is inseparable from the decision of which data to use. By using only the parallax (and its uncertainty), we are free from the influence of a Galactic model, and can provide distance estimates where there is no colour information. Yet this freedom (or voluntary ignorance) reduces precision. By allowing a Galaxy model to influence our results to some degree – either weakly via $L(l, b)$ as in paper IV, or more strongly as in the 3D MW model of paper III – we can improve precision. If we are willing to adopt further assumptions, such as an HRD, we can also introduce more data, such as the magnitude and colour (as described in section 4 of paper III, for example), thereby increasing the precision (and hopefully the accuracy) further.

I now provide some ideas on how the distance estimates could be improved in the future.

1. The EDSD prior is quite asymmetric. It has considerable probability at small distances ($r < 2L$), but arbitrarily low probability density at large distances ($r \gg 2L$). Thus for large fpu, large distances are suppressed more than small ones. Given the underestimation issue (item 1 on page 5) and the apparent lack of a significant overestimation issue (although this would need to be investigated properly), one remedy would be to use larger values of L. This would have little influence on the inferred distances when the fpu is small, but could lead to a higher accuracy when the fpu is large (due to the assumption that such stars are generally further away than the prior in paper IV implies). In other words, we could tune the prior not for the set of observed stars as a whole, but for the set of stars which would benefit more from it.

2. One way of making the prior work more where we need it it to use a prior that depends not only on (l, b), but also[2] on σ_ϖ. Figure 2 shows, using the GDR2 mock catalogue (Rybizki et al. 2018), the dependence of distance on σ_ϖ. Or rather it shows a lack of useful dependence. Conceptually we anyway seem to want a prior that depends on the parallax SNR, $s = \sigma_\varpi/\varpi$, rather than σ_ϖ. Figure 3 shows that we do indeed see some dependence of distance on s. However, because ϖ is a noisy quantity for which we already have a likelihood, a "prior" that depends on parallax SNR is not a prior: we cannot combine $P(r \mid \sigma_\varpi/\varpi)$ with the likelihood $P(r \mid \varpi)$ to get the posterior $P(r \mid \varpi, \sigma_\varpi)$. There appears to be no meaningful way to additionally use anything that depends on the parallax.

   Figures 4 and Figures 5 do, however, show that there is a dependence of distance on G and $\mathrm{BP} - \mathrm{RP}$. (Recall that σ_ϖ depends on these, but also on the number of observations.) If we consider these as effectively noise-free (in the same way that we treat l,b as noise-free) then we could consider extending the prior to be a four-dimensional function $P(r \mid l, b, G, \mathrm{BP} - \mathrm{RP})$, and construct this from the mock catalogue. Alternatively, we introduce an additional likelihood on G, and connect this to the distance via the absolute magnitude M_G and extinction. We then place a prior on M_G (e.g. constructed from the mock catalogue) and marginalize over it. We take this approach, albeit formulated in a slightly different way, in section 9.

3. Both the likelihood and the EDSD prior are unimodal in r, so their product is in general bimodal. Although this only occurred for 0.09% of sources in GDR2 (see paper IV), the very fact that it can occur complicates the general approach to summarizing the posterior. I consider a simplified version of this prior that gives a unimodal posterior in section 7.

[2] Given that in the simple distance inference (e.g. papers I and IV) the likelihood is $P(\varpi \mid r, \sigma_\varpi)$ and the posterior is $P(r \mid \varpi, \sigma_\varpi)$, Bayes' theorem shows us that the prior is formally $P(r \mid \sigma_\varpi)$ and not just $P(r)$. In the past we defined a prior independent of σ_ϖ.

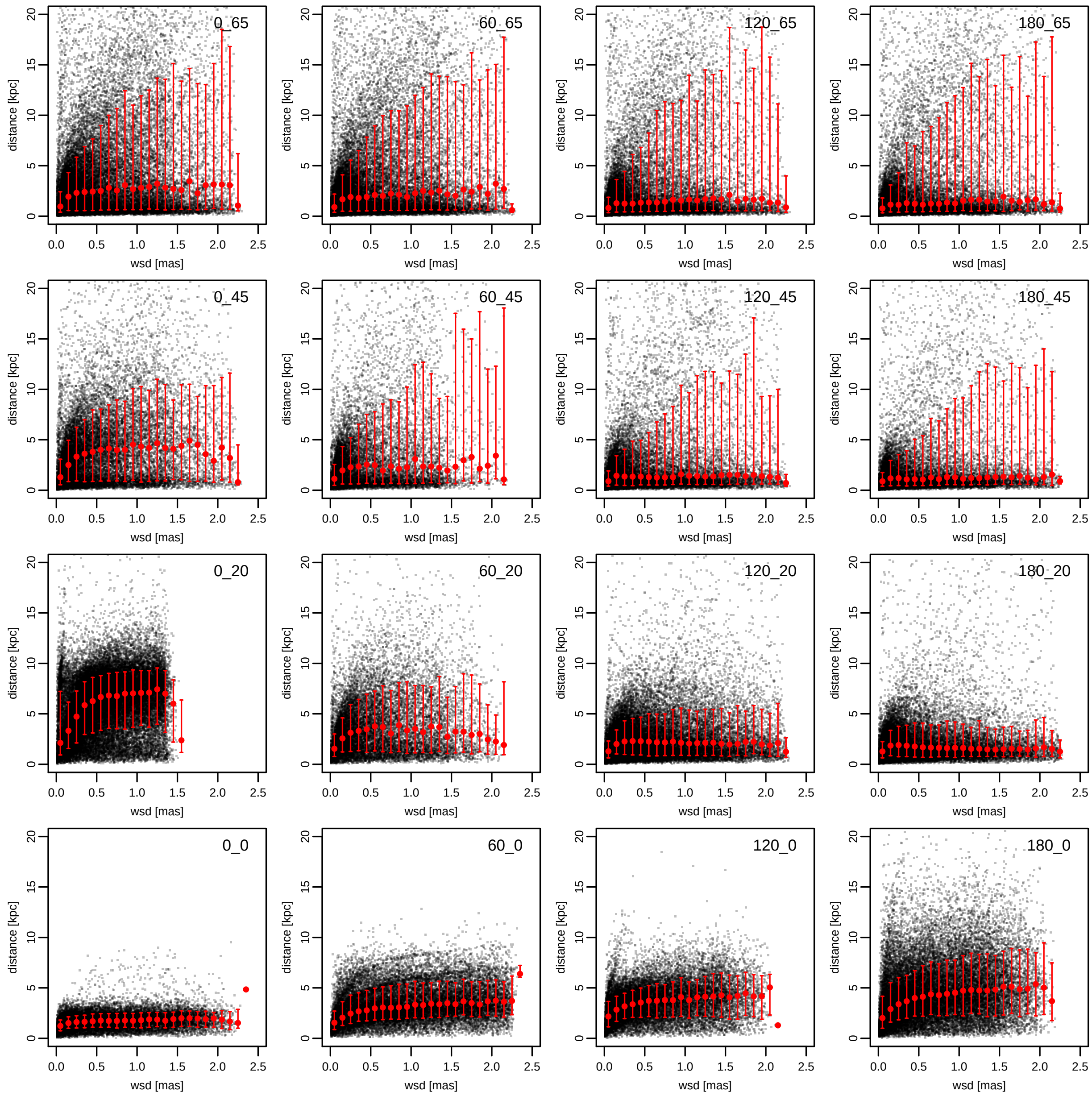

Figure 2: The variation of true distance (vertical axis) with σ_ϖ (horizontal axis) for various lines-of-sight indicated by l_b in the upper right corner of each panel, computed from the GDR2 mock catalogue (Rybizki et al. 2018). The black open circles are individual stars (each line-of-sight is between 0.3 and 2.0 degree in radius). The red point are the median and the error bars the 16% and 84% quantiles of the distribution in contiguous SNR bins of width 5.

4. One of the limitations of the EDSD prior is that it has only one free parameter, the length scale L. Writing it as $r^\beta \exp(-r/L)$ for $\beta \geq -1$ and $L > 0$, this has a mean of $(\beta+1)L$ and a standard deviation of $\sqrt{\beta+1}L$ (and a mode of βL provided $\beta \geq 0$). So if we are change β from 2 (the value that gives the EDSD a constant volume density for $r \ll L$), then we can vary both the shape and the scale of the prior distribution. Specifically, we could then achieve a narrower or broader prior (for given length scale) to reflect a wider

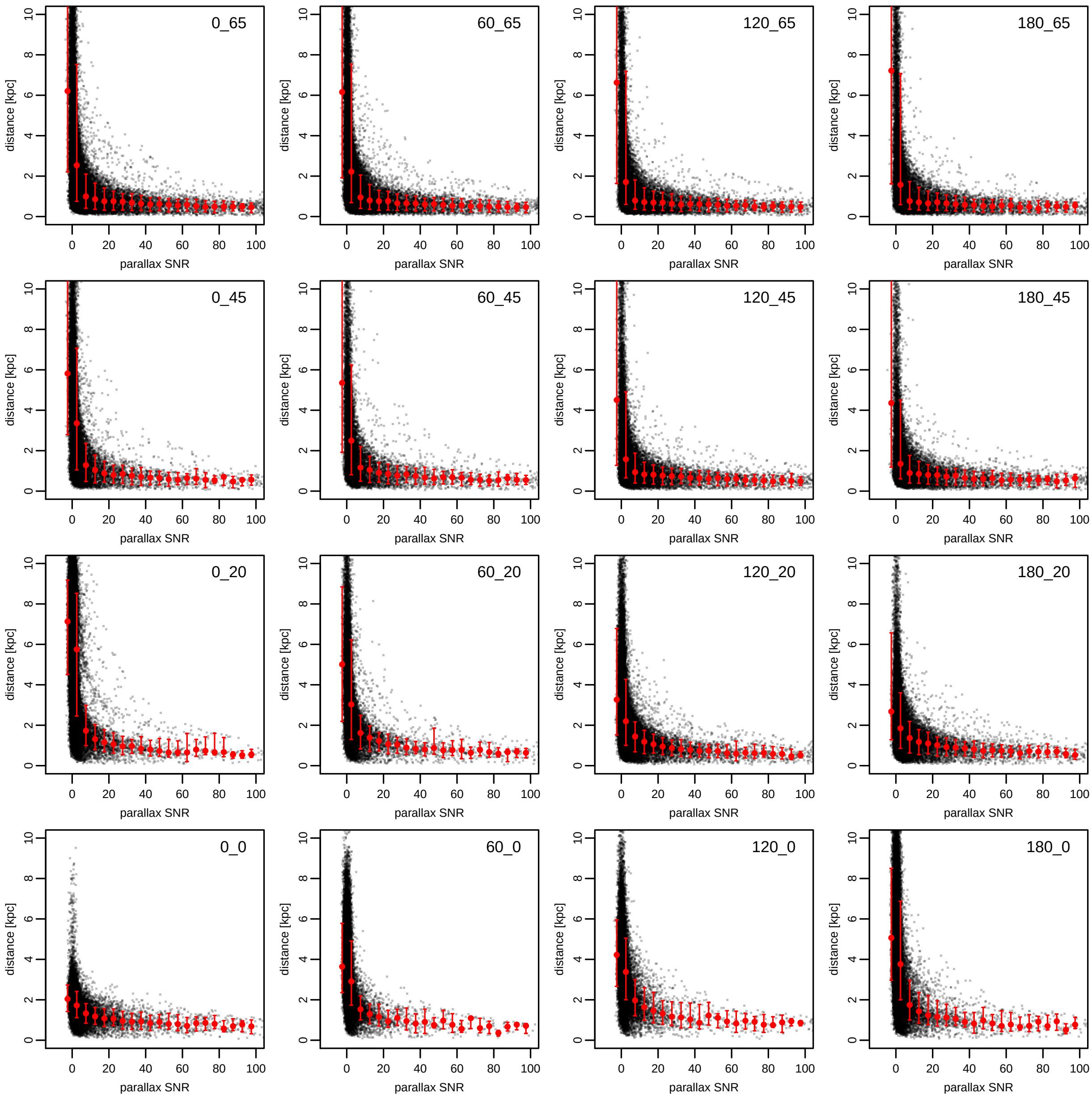

Figure 3: As Figure 2, but now showing the variation against the parallax SNR, which has been computed using a noise-perturbed parallax (i.e. this is a measureable quantity).

a priori spread of distances. This could be used in a prior conditioned on G, for example, as we see in Figure 4 a strong dependence of the distance spread on G. I explore a generalization of the EDSD prior in section 8.

5. There is no theory that says which estimator – e.g. mean, median, or mode – is optimal for a given distribution, let alone for a given type of distribution.[3] The mode was adopted in paper IV because it seems natural for the asymmetric distance posterior, plus the posterior

[3] Despite some apparent claims, there is no relevant asymptotic limit here, because the data are what they are. In the limit $\sigma_\varpi \rightarrow 0$ any non-pathological posterior becomes a delta function at $1/\varpi$, so all estimators are identical.

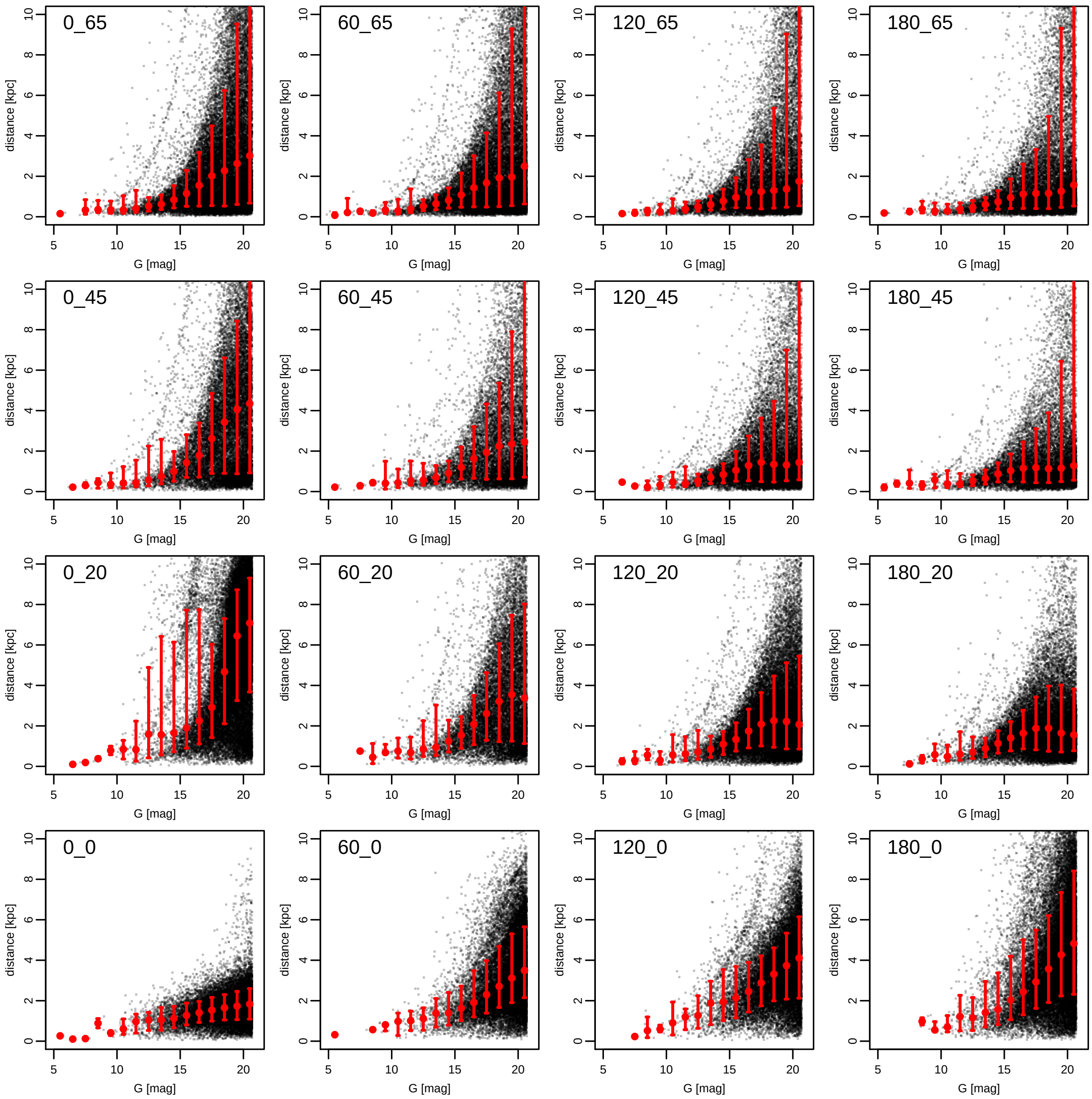

Figure 4: As Figure 2, but now showing the variation against G.

arising from the EDSD prior is quick to compute (it is the solution of a cubic equation). The HDI had to be computed numerically, but typically involved just 100 evaluations of the posterior. (Note that computation of the HDI requires normalization of the posterior; we did this using Gaussian quadrature.) Yet the mode does not tranform invariantly (see item 2 on page 5). The median does, but its computation involves integration or sampling (the former has to be done numerically for any sensible priors). If we reported the median it would make more sense to report an ETI than an HDI – the median does not necessarily lie within the HDI – and computation of the ETI also requires integration or sampling. In paper IV we used MCMC for that.

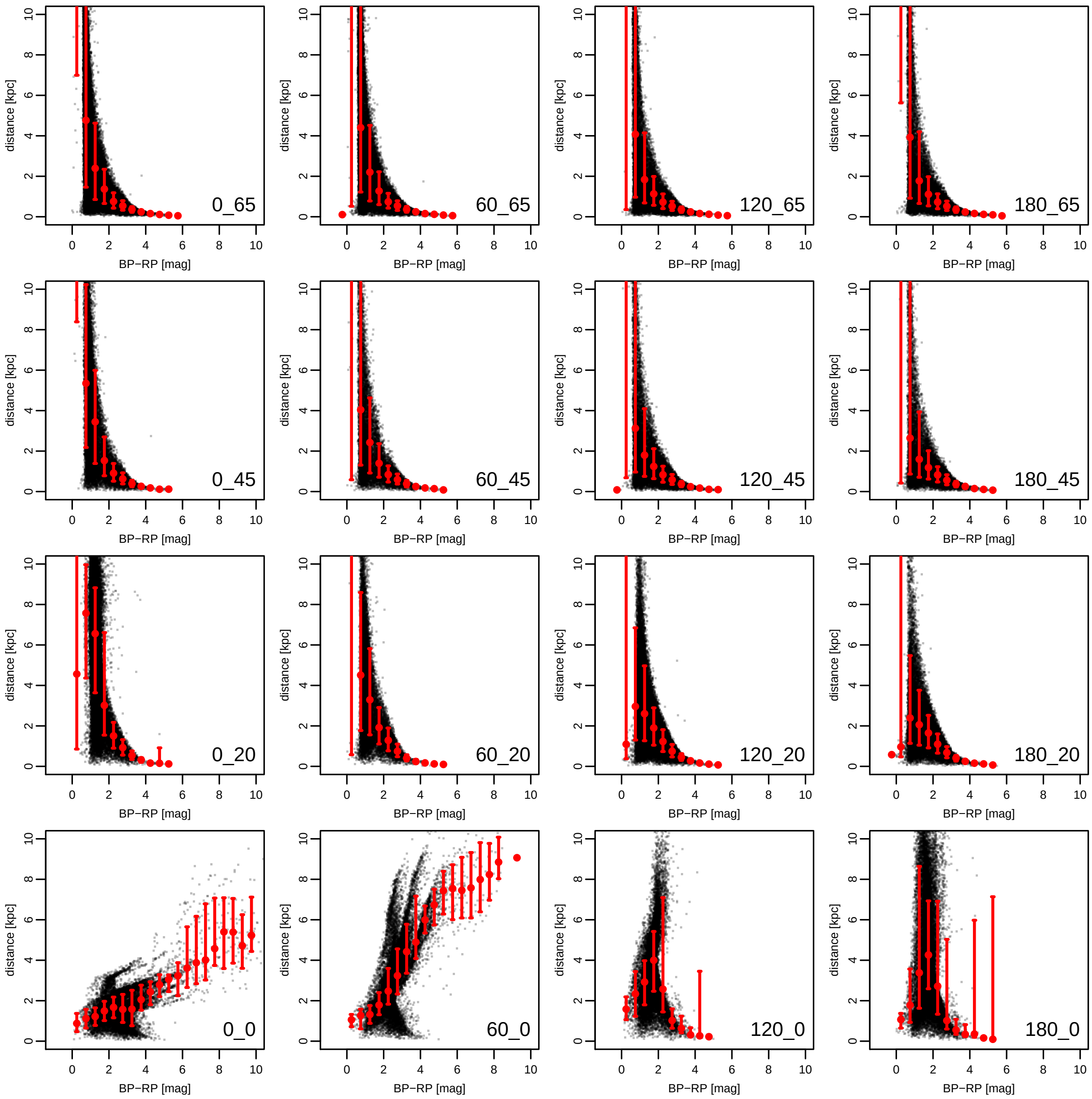

Figure 5: As Figure 2, but now showing the variation against $\mathrm{BP}-\mathrm{RP}$.

6. Some users wanted to do better than invert the parallax, but didn't want to use our length scale model from paper IV. Hence there is some value in providing a tool to enable users to specify their own value of L. One could extend this to offer multiple forms for the prior, and/or different summaries of the posterior. Such a tool could take as input a list of Gaia source IDs (and the choice of prior and posterior summaries) and interrogate the online Gaia catalogue in real time.

7. Some users had additional information that they wanted to include in the distance inference. This is a very open-ended use case, involving many different possible additional assumptions and models. One convenient way to express additional information is in the

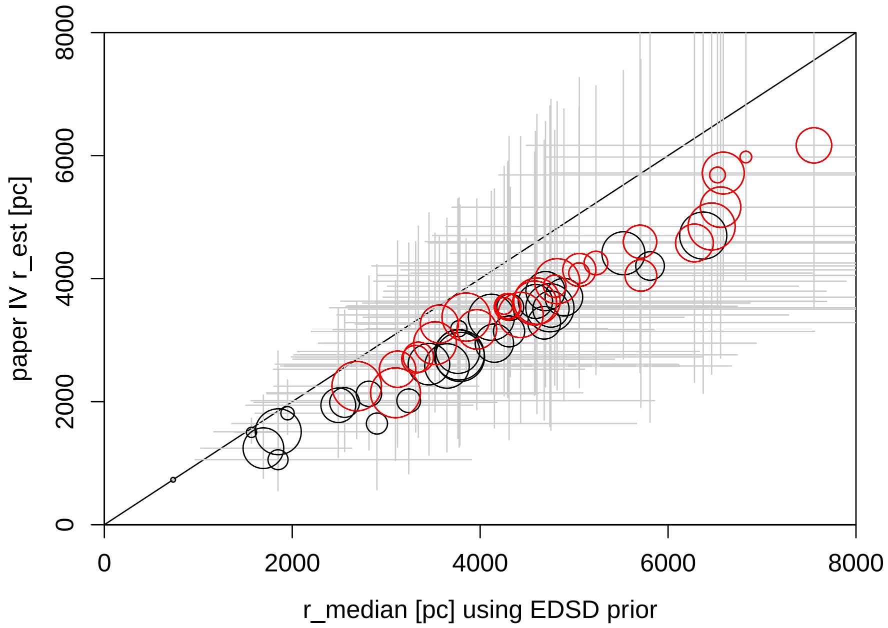

Figure 6: Comparison of median and mode estimators using the EDSD prior for stars with fpu ranging from -3 to $+3$ for stars with a median distance of less than 8000 pc. (The vertical axis is equal to the estimate from paper IV.) The area of each circle is proportional to the magnitude of fpu, with positive values show in black and negative values shown in red. The diagonal black line is the identity.

form of a prior on M_G, $P(M_G)$. One way to introduce this is developed in section 9.

5 Median and equal-tailed confidence interval as estimators with the EDSD prior

In paper IV we used the mode and 68% HDI to summarize the posterior. We compare these estimates to the median and 68% ETI for 60 stars covering a range of fpu in Figure 6. Figure 7 shows the ratio of the point estimates (median/mode) for a larger sample of stars as a function of fpu (without any distance limit). We see that the median is never smaller than the mode, and that for small positive fpu up to about 0.2 it is not much larger than the mode. Figure 8 shows a similar comparison but for the corresponding confidence intervals, i.e. the ETI (fr the median) and the HDI (for the mode). We see that the upper and lower bounds, as well as the width of the interval, tend to be slightly larger for the HDI.[4] This initial suggests there is nothing particular adverse about using the median and ETI as estimators instead of the mode and HDI.

[4]For a unimodal posterior, the width of the HDI should never be less than the width of the ETI, because the HDI is then smallest intervals. But some of these posteriors are bimodal, plus the method of computing the HDI in paper IV can lead to it being slightly larger than the 68.1% interval.

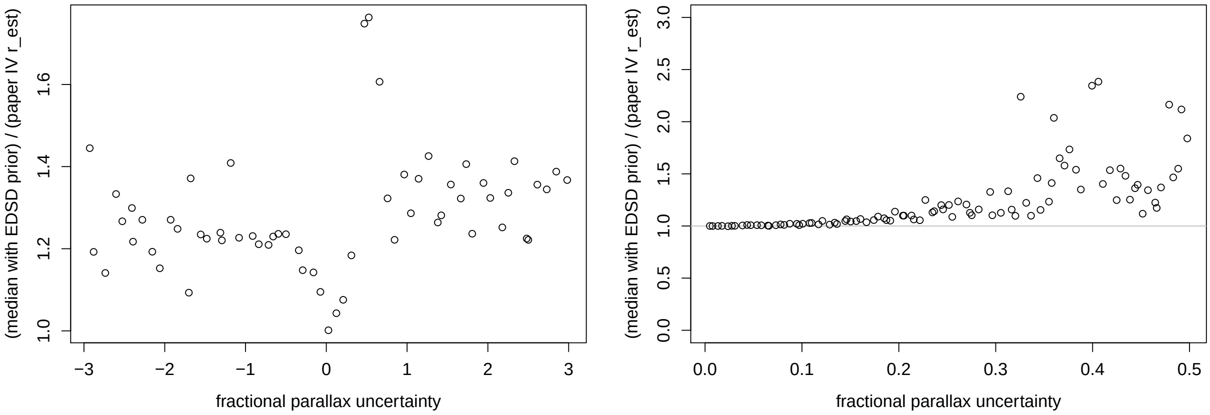

Figure 7: Ratio of the median to the mode of the posterior using the EDSD prior (the latter being the estimate in paper IV), as a function of fpu for various stars from GDR2. The right panel shows more stars over a narrower range of fpu (a few points extend beyond the top of the plot in this panel).

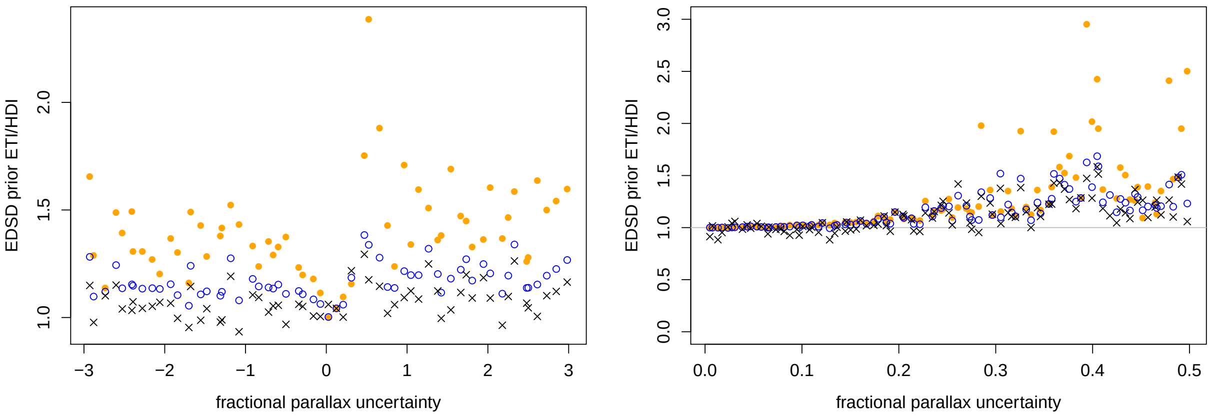

Figure 8: Ratio of measures of the ETI to measures of the HDI for the EDSD prior, as a function of fpu for various stars from GDR2 (the HDI is the uncertainty estimate used in paper IV). Orange closed circle: ratio of the upper bounds (84.1-percentile); blue open circle: ratio of the lower bounds (15.9-percentile); black cross: ratio of the widths of the confidence intervals. The right panel shows more stars over a narrower range of fpu (a few points extend beyond the top of the plot in this panel).

6 A more accurate inference of the EDSD prior length scale

In paper IV we computed the length scale of the EDSD prior, L, from the distances $\{r_i\}$ to the stars in the mock catalogue in each HEALpixels. Theoretically $L = \overline{r_i}/3$, but we replaced $\overline{r_i}$ with the median, on the grounds that it is less influenced by a few stars at very large distances. This introduces a small theoretical error, however, because the median is always smaller than the mean for the EDSD prior. This can be seen in Figure 9, which plots the median vs the mean ($= 3L$) of the prior. Although there is no closed analytic solution for the median, it is

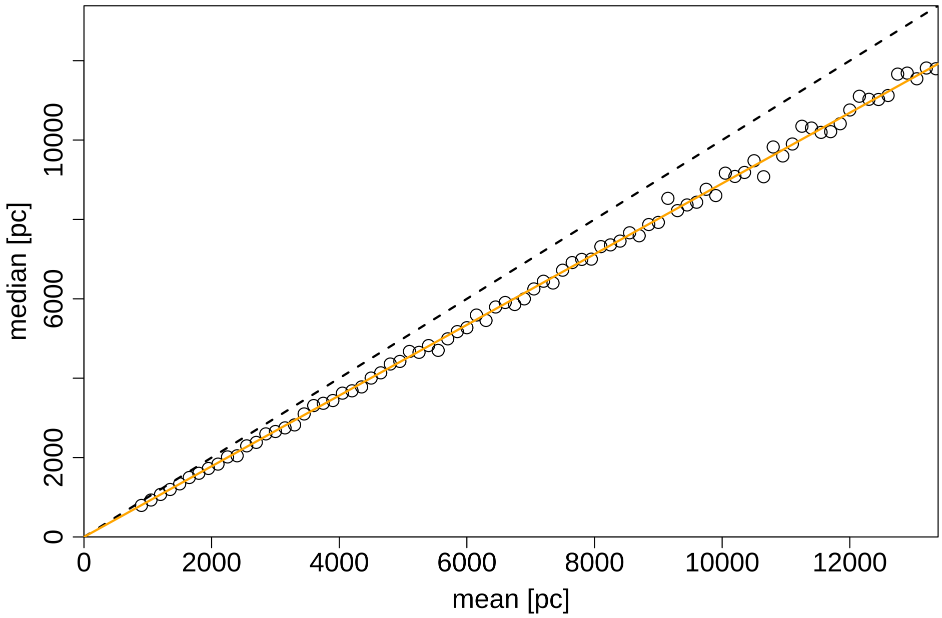

Figure 9: Relationship between the median and mean ($= L/3$) of the EDSD prior for a range of length scales. The dashed line shows the one-to-one line. The orange line is the least squares fit with intercept constrained to zero; it has a slope of 0.892.

easily computed via MCMC (here using 10^4 samples). We see a close linear relationship over a wide range of L. The slope of the least squares fit (with intercept fixed to zero) is 0.892, i.e. the median is 12% smaller than the mean. Equivalently the median is $2.676\,L$. Thus by having used the median of the distances from the mock catalogue, we were underestimating L by about 12%. It would therefore be better to compute L from the mock catalogue as $L = 0.374\,r_{\rm median}$ (rather than $L = r_{\rm median}/3$, as done in paper IV).

7 Exponentially decreasing distance (EDD) prior

This prior is defined as

$$P(r \mid L) \;=\; \begin{cases} \dfrac{1}{L}\,e^{-r/L} & \text{if } \; r > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

and so the unnormalized posterior is (for $L > 0, \sigma_\varpi > 0$)

$$P^*(r \mid \varpi, \sigma_\varpi, L) \;=\; \begin{cases} \exp\left[-\dfrac{r}{L} - \dfrac{1}{2\sigma_\varpi^2}\left(\varpi - \dfrac{1}{r}\right)^2\right] & \text{if } \; r > 0 \\ 0 & \text{otherwise}\,. \end{cases} \tag{4}$$

where parallax may be adjusted for a zeropoint, uncertainties could be modified, and L could depend on other measured quantities. The prior is unimodal, at $r = 0$. The motivation for this

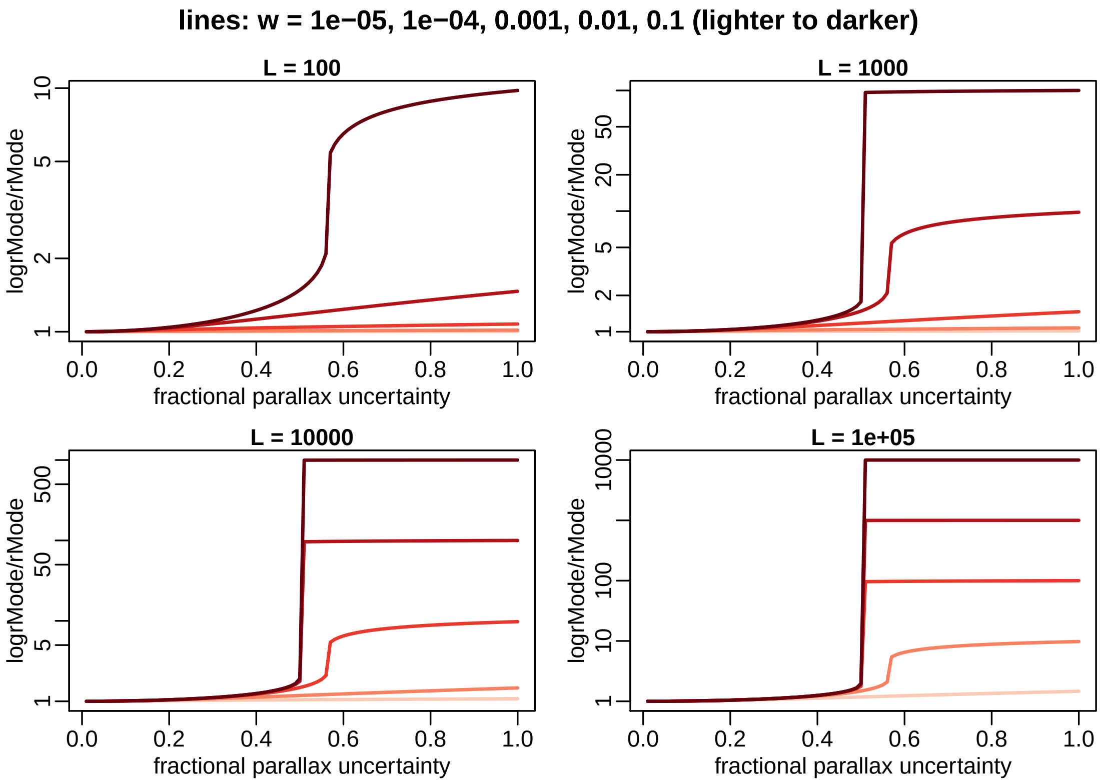

Figure 10: As Figure 1, but now for the EDD prior.

prior is that the posterior is also unimodal, with $r_{\rm mode}$ given by the solution of the cubic equation

$$\frac{r^3}{L} + \frac{\varpi}{\sigma_\varpi^2} r - \frac{1}{\sigma_\varpi^2} = 0 \,. \tag{5}$$

This is the same as the solution for the EDSD prior (equation 19 of paper I), except that now the r^2 coefficient is zero. This too has three complex roots in general, but given the positivity of L and σ_ϖ, only one real positive root in practice.

The space density, $P(V)$, for any prior is given by $P(V) \propto P(r)/r^2$. For the EDD prior, $P(r \to 0) \to 1/r^2 \to \infty$, which is unphysical. In practice this may not matter too much, however, as for most nearby stars the likelihood dominates in the posterior. The EDSD prior can be seen as a modification of the EDD prior to avoid divergence, as for the EDSD prior $P(r \to 0) \to \rm const.$

As already discussed (item 2 on page 5), the mode of the posterior over the log distance[5] (using the same prior) is not $\log(r_{\rm mode})$. Any hopes that these two modes may not be so different for this posterior are dashed when we look at Figure 10. If we want to have a single number which can be used as the best estimate for both distance and (when logged) distance modulus, then we must instead use a quantile, and this requires integration or sampling (both of which

[5]The mode of $\log r$ is given by equation 5 with an additional $-r^2$ term in it.

must be done numerically). Although the posterior in equation 4 is not a complicated function, determining its quantiles quickly (i.e. with a small number of evaluations of the posterior) is imperative, as we need to do it for 10^9 sources within a few days. (For comparison, computing the distances for the 1.33 billion stars in paper IV took 75 CPU-core-days.)

The most straight-forward approach to computing quantiles is with MCMC.[6] If we use the Metropolis algorithm, then it seem sensible to initialize the sampling at the mode. We also need a step size. The characteristic width of the posterior varies enormously, from 10^{-4} pc for the nearest stars to 10^4 pc for stars with negligible parallaxes (as this is the width of the prior for largest L). A suitable step size, s, can be obtained by approximating the log posterior as a Gaussian about its mode, the variance of which is

$$s^2 = \left(- \frac{\mathrm{d}^2 \ln P^*}{\mathrm{d}r^2} \Big|_{r_{\mathrm{mode}}} \right)^{-1} \tag{6}$$

$$= \frac{\sigma_\varpi^2 r_{\mathrm{mode}}^4}{3 - 2\varpi r_{\mathrm{mode}}} \, . \tag{7}$$

As the second derivative must be negative at the mode, this expression also tells us that, for this posterior, the mode is always less than 1.5 times the inverse of the parallax (for positive parallaxes).

How should we choose L? For the EDSD prior in paper IV we took a Galaxy model, binned it into small contiguous cells of (l, b), then maximized the product of the priors for all the stars along that line-of-sight. This gave $L = \frac{1}{3}\bar{r}$, where $\bar{r}$ is the mean of the distances (although in fact we used the median as a more robust estimate in order to exclude a few stars at very large distances in the model; see section 5.5 of CBJ-080). The same principle applied to the EDD prior gives $L = \bar{r}$ (and again we could replace this with the median).

The posteriors for 30 stars from GDR2 using this prior and length scale are shown in Figures 11 and 12 for fpu increasing from -1 to $+2$. (In this and all plots which follow, sources have been selected to have an approximately uniform distribution in fpu.) I sampled these posteriors using the Metropolis algorithm for 2000 iterations (after 100 burn-in steps) initialized at r_{mode} with a step size equal to s (equation 7). From the resulting samples I computed the median, r_{med}, as well as the 0.159 and 0.841 quantiles which together form the 68.2% equal-tailed confidence interval (ETI). We see from Figures 11 and 12 that the sampling is reasonably successful, and the quantile summaries look plausible. A comparison of the median to the mode of the posterior for a wider range of fpu is shown in Figure 13. As expected, this ratio is close to one for small fpu. For larger fpu and negative parallaxes this ratio can be 2 or more. A comparison of the posterior median to the distance estimate from paper IV (the mode using the EDSD prior) is shown in Figure 14. Their ratio is also close to one for small fpu. For larger values of fpu and negative parallaxes, the EDD median is generally larger than the EDSD mode, by up to 70% for the fpu range shown. Figure 15 shows an alternative comparison of these distances, just for nearer stars. A comparison with the naive inverse parallax distance (just for positive parallaxes) is shown in Figure 16. This may be compared to Figure 6 (panels h and i) in paper IV.

[6] There don't appear to be any obvious shortcuts involving splitting the function into the low r and high r extremes and the core, and approximating each using error functions and/or Gaussian quadrature.

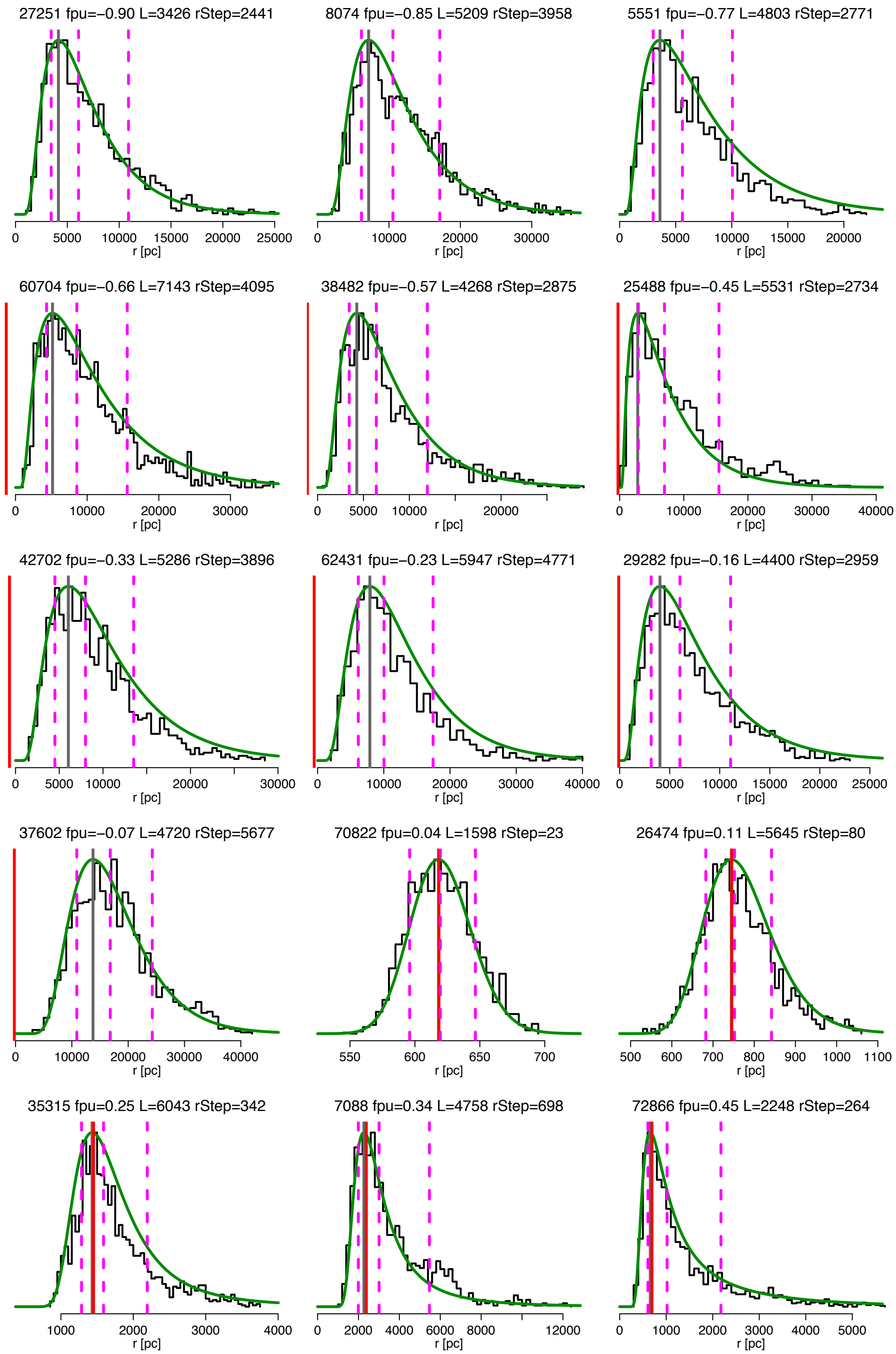

Figure 11: Examples of posteriors computed using the EDD prior for a range of fpu. Green is the posterior, the black histogram is the MCMC sampling thereof (scaled to have the same maximum). The vertical magenta lines show the 0.159, 0.5, and 0.841 quantiles. The vertical red line is $1/\varpi$. The vertical grey line is the posterior mode. `rStep` is s from equation 7.

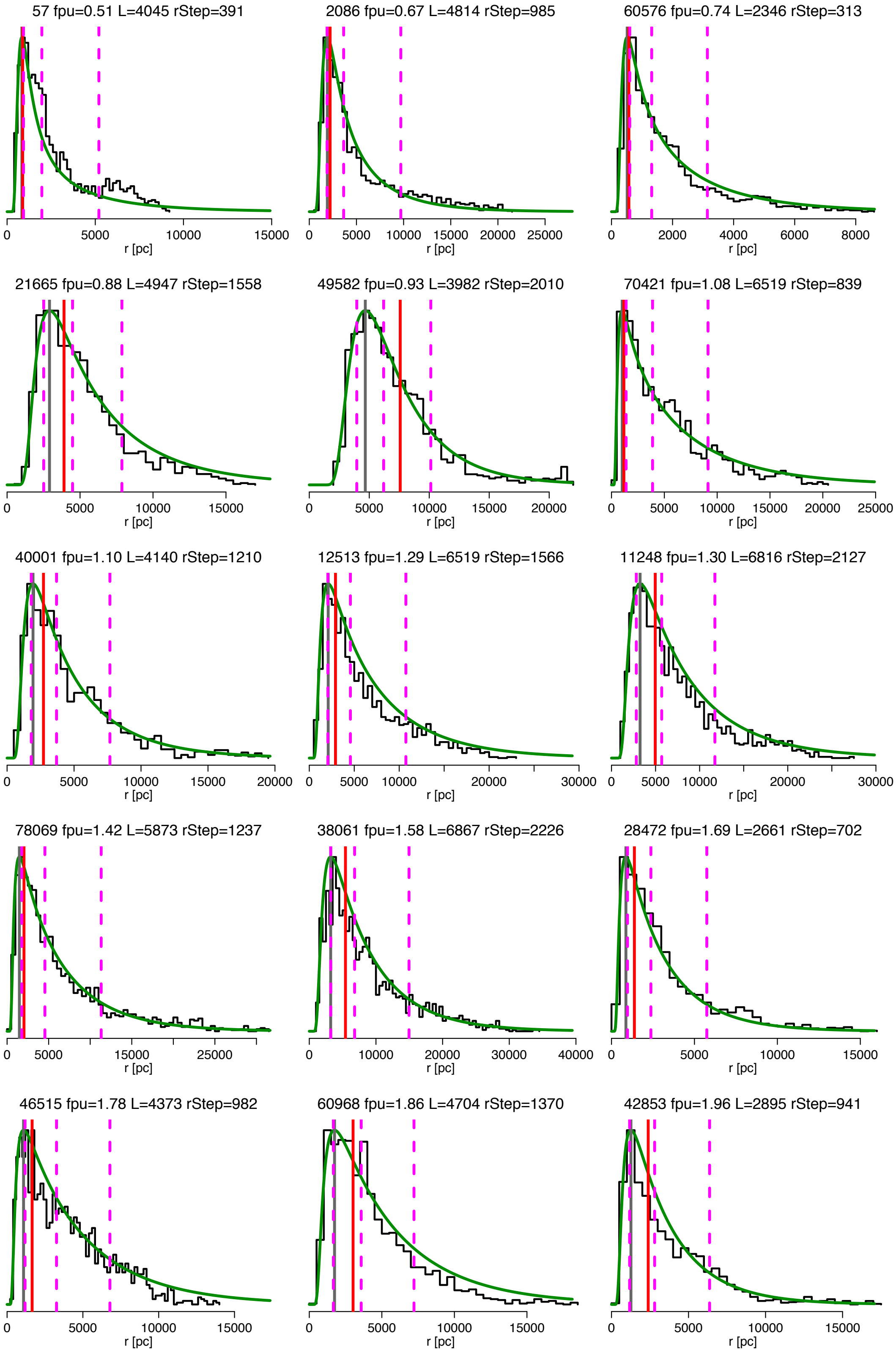

Figure 12: Continuation of Figure 11.

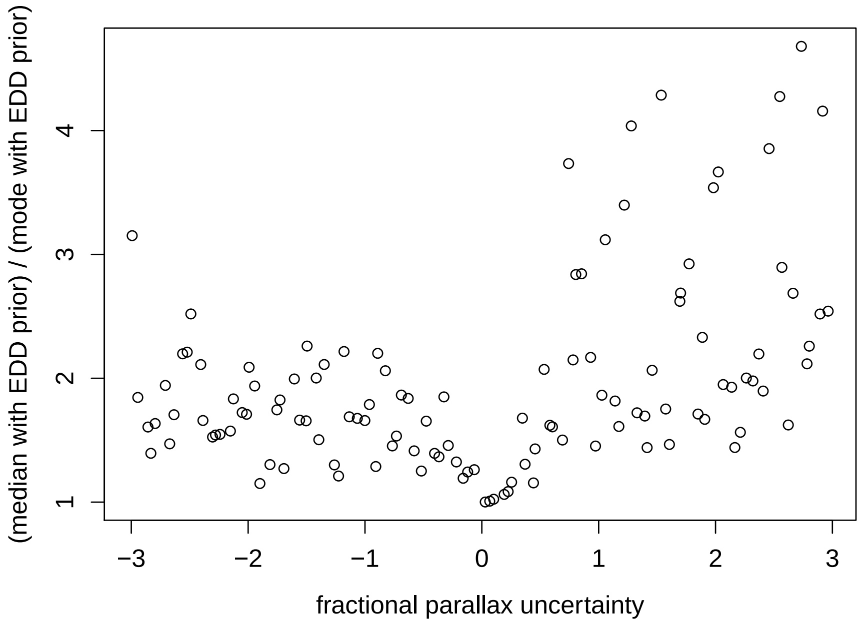

Figure 13: Ratio of the median to the mode of the posterior using the EDD prior, as a function of fpu for various stars from GDR2.

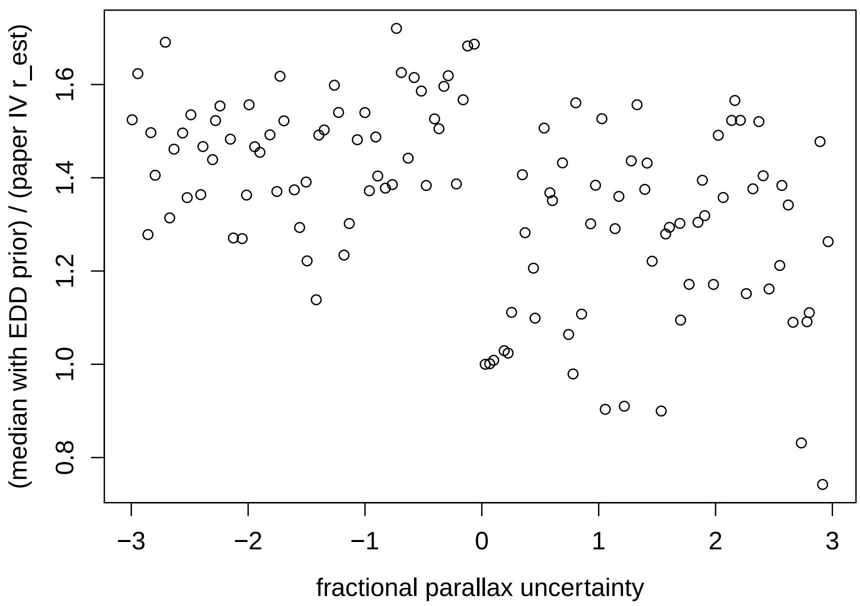

Figure 14: Ratio of the median using the EDD prior (computed here) to the mode using the EDSD prior (from paper IV), as a function of fpu for various stars from GDR2.

My laptop took 0.0275 seconds to compute the quantiles in the above way for one star (computed from 10 000 stars selected at random from GDR2). Thus 1.33 billion stars would require about 425 CPU-core-days. This is only about six times longer than needed for the paper IV catalogue (which makes sense, as this is roughly the ratio of the number of posterior evaluations used by the two methods). No effort has yet been expended, however, to determine a sufficient

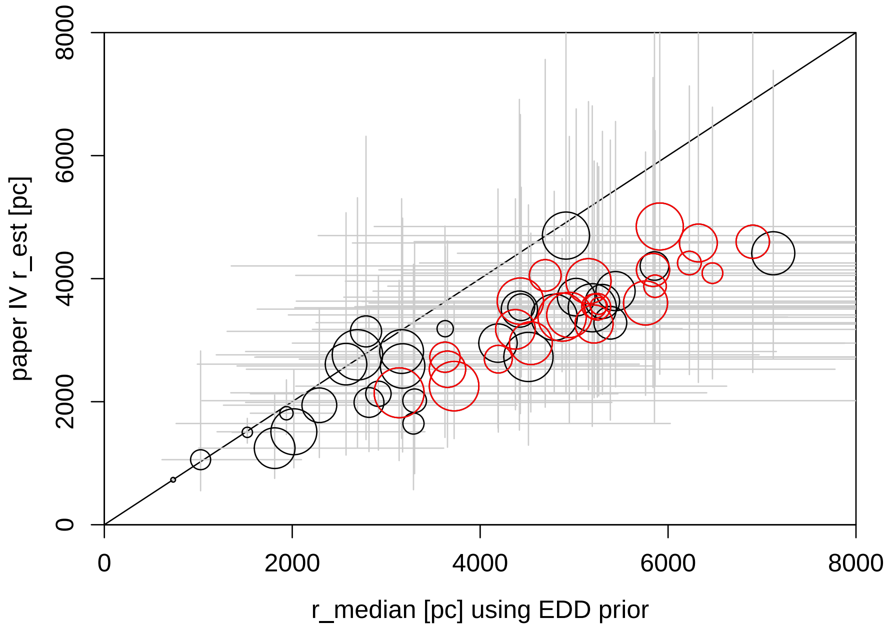

Figure 15: Comparison of median using the EDD prior (computed here) to the mode using the EDSD prior (from paper IV) for stars with fpu ranging from -3 to $+3$ for stars with a median distance of less than 8000 pc. The area of each circle is proportional to the magnitude of fpu, with positive values show in black and negative values shown in red. The diagonal black line is the identity. These are the same stars – and same quantity on the vertical axis – as in Figure 6.

yet minimal number of samples.

8 Generalized gamma distribution (GGD) distance prior

Both the EDSD and EDD priors are special forms of the Generalized Gamma Distribution (GGD), which can be written (as a distance prior) as

$$P(r \mid L, \alpha, \beta) \;=\; \begin{cases} \dfrac{1}{\Gamma(\frac{\beta+1}{\alpha})} \dfrac{\alpha}{L^{\beta+1}}\, r^{\beta}\, e^{-(r/L)^{\alpha}} & \text{if} \;\; r > 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

for $\alpha > 0$, $\beta > -1$, and $L > 0$. $\Gamma()$ is the gamma function. The mode of this distribution is $L(\beta/\alpha)^{1/\alpha}$ for $\beta > 0$, and at zero otherwise. The mean is $L\,\Gamma((\beta+2)/\alpha)/\Gamma((\beta+1)/\alpha)$. Special cases are: EDSD prior $\alpha = 1, \beta = 2$; EDD prior $\alpha = 1, \beta = 0$; gamma distribution $\alpha = 1$. Note that with $\alpha = 1$, $\beta > 2$ means the volume density of stars initially increases with distance from the Sun, and $\beta < 2$ means it decreases.

Figure 17 shows some examples of this prior. The two panels in the middle row show the effect of varying α for $\beta = 2$. When $\alpha = 1$ we have the EDSD prior. We expect that decreasing α below 1 will make the function decay less with increasing distance. We observe this, but also

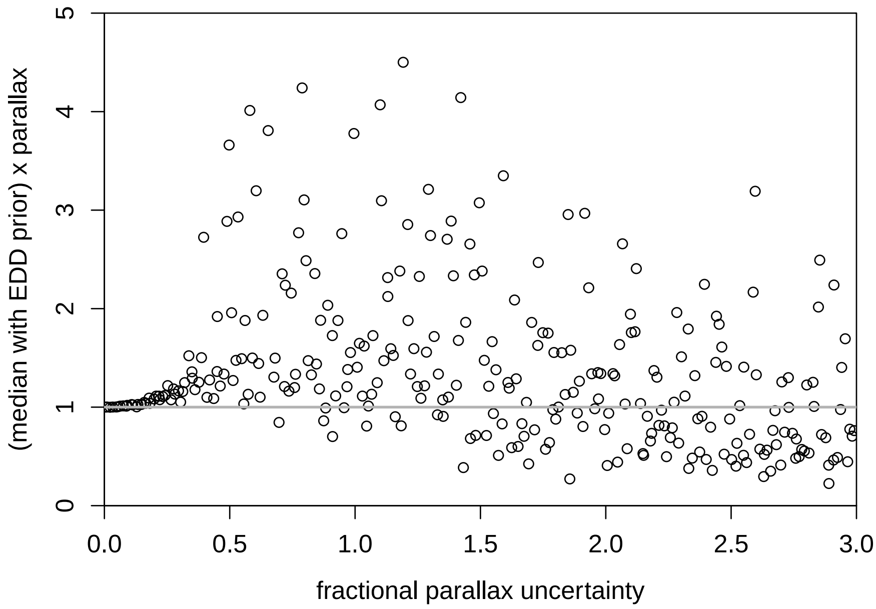

Figure 16: Ratio of median distance using the EDD prior (computed here) to the inverse parallax distance for 300 stars with fpu ranging from 0 to $+3$. The horizontal grey line is the identity. Four stars have $r_{\rm med}\varpi > 5$ and so are omitted from the plot.

a shift in the mode to larger distances. Decreasing α even by small amounts flattens out the function quite quickly. Moreover, to keep the mode to less than a few thousand parsec, we need to decrease the length scale (L) significantly as α drops below about 0.7 (middle row, left panel). This can be seen better in the top row of the figure, where the two panels show the variation of the distribution with L for two different small values of α: go get useful distributions over a range of a few thousand parsec, we need values of L of just a few parsec. If we set $L = 1000$ for $\alpha \leq 0.6$, then the mode is at $r \geq 7438$ pc. The shape of this prior is very sensitive to α, and is the impact of α is quite dependent on the value of L. Thus while varying α in the GGD prior gives a prior with more flexibility than the EDSD prior, we can no longer interpret L as a physically-meaning length scale. The parameter β too has quite some impact, and varying it also changes the interpretation of L, as we can see in the bottom two panels. Recall that $\beta = 2$ implies a uniform space density of sources as $r \to 0$: $\beta > 2$ therefore implies that the density initially increases with distances (although of course the exponential decay gives a decrease in density sooner or later).

Figure 18 shows the result of fitting the GGD prior in all three parameters to data from the GDR2 mock catalogue of Rybizki et al. 2018. Each panel shows the density distribution of the data (in black) for a given HEALpixel (at level 4) indicated by p. Each orange lines shows the fit of the GGD prior (fit by maximizing the likelihood the mock data). Figure 19 shows the same data, but as cumulative distributions, and Figure 20 shows the ratio of the fits to the data. The fits looks quite good in most cases. The parameters of the fit are given in the orange text. We see that α is often bewteen 0.2 and 0.3, and never near to 1.0, and L is often much less than 1 pc. For comparison I also fit the EDSD prior, where "fit" just means setting L to 0.374 times the

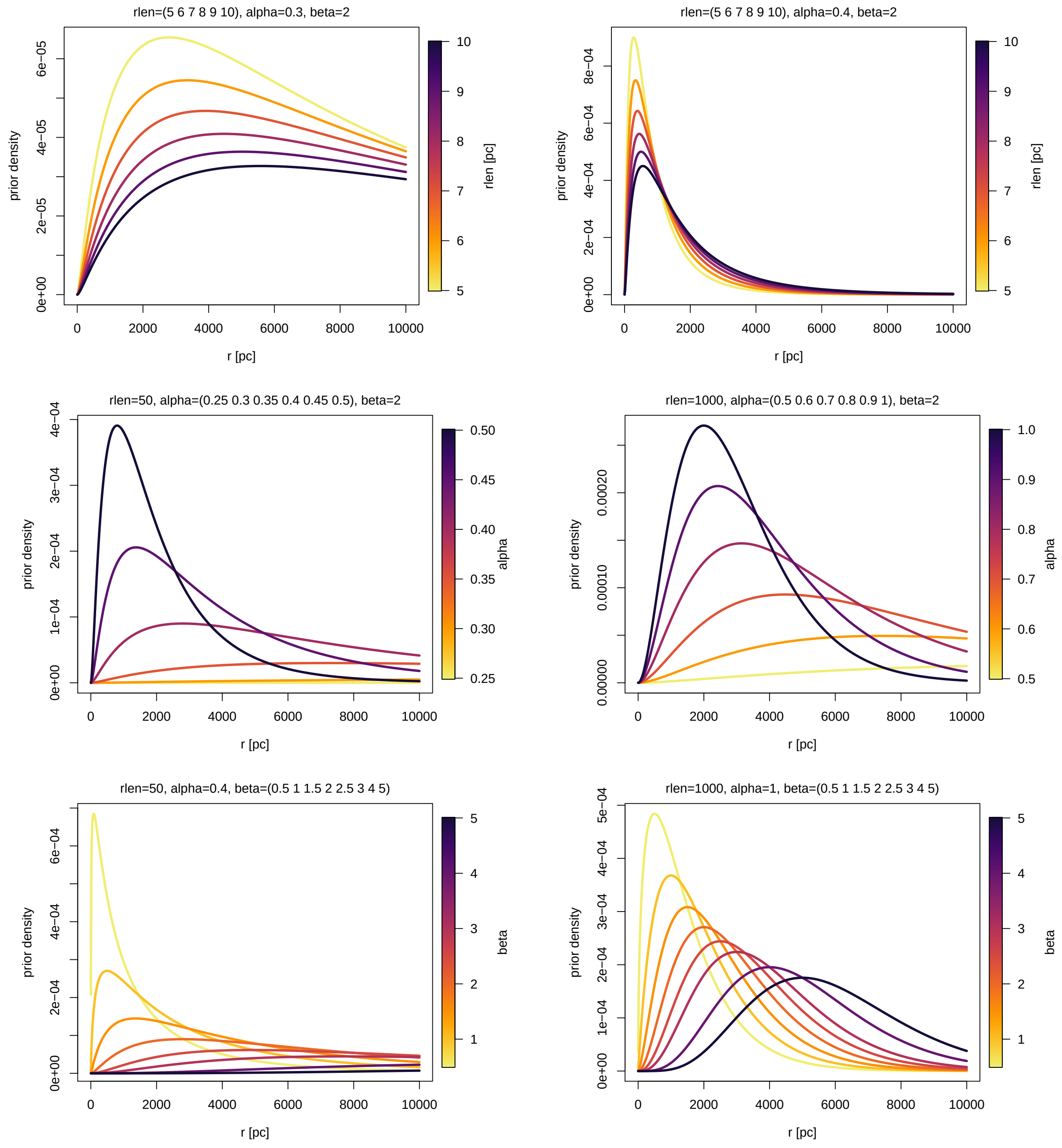

Figure 17: The generalized gamma distribution, as a potential distance prior, for various values of its parameters. In all cases it is plotted over the distance range 0–10 000 pc. Top panels: variation in L ("rlen") for $\alpha = 0.3$ (left) and $\alpha = 0.4$ (right); Middle panels: variation in α (two ranges) for $L = 50$ (left) and $L = 1000$ (right); Bottom panels: variation in β for $L = 50, \alpha = 0.4$ (left) and $L = 1000, \alpha = 1$ (right).

median of the data (see section 6). This is plotted as the dashed blue line. These fitted values of L, written in blue, are of course much larger. Visually, the GGD prior fits the data better than the EDSD prior in every case, and this is confirmed by a comparison of the Kolmogorov–Smirnov statistics (the maximum distance between the cumulative distributions of the data and the fit). However, the EDSD fit looks reasonable in quite a few cases.

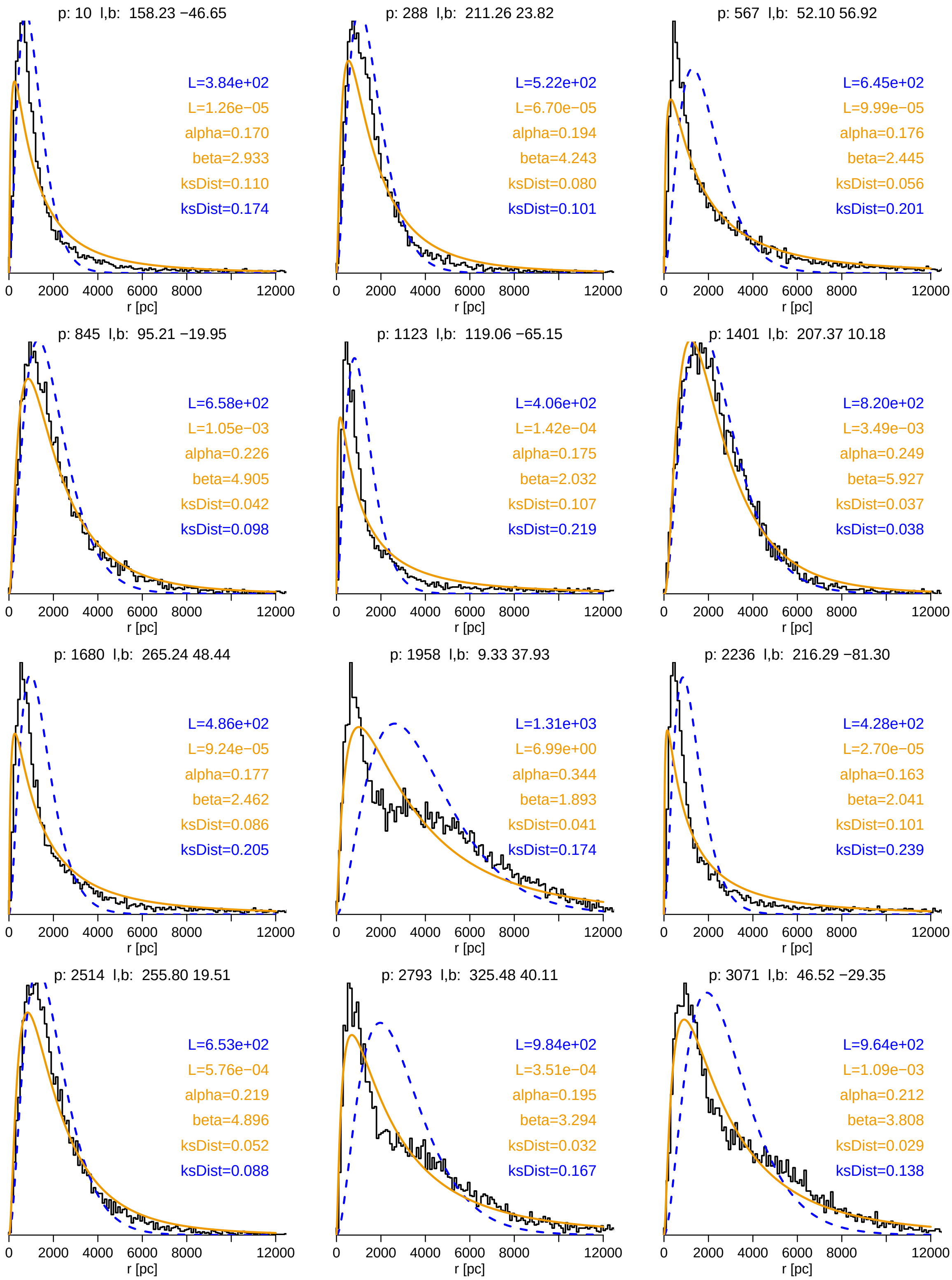

Figure 18: Fits of the GGD prior (orange solid lines) and EDSD prior (blue dashed line) to the distances to stars (black histogram) in 12 different HEALpixels (panels) selected from the GDR2 mock catalogue. All distributions are normalized, and the vertical scale is linear (scaled to the range of the histogram). p indicates the HEALpixel number (level 4), l and b are the average Galactic longitude and latitude of the stars in it. The coloured text shows the parameters of the fits. ksDist is the KS-test distance between the data and the fit.

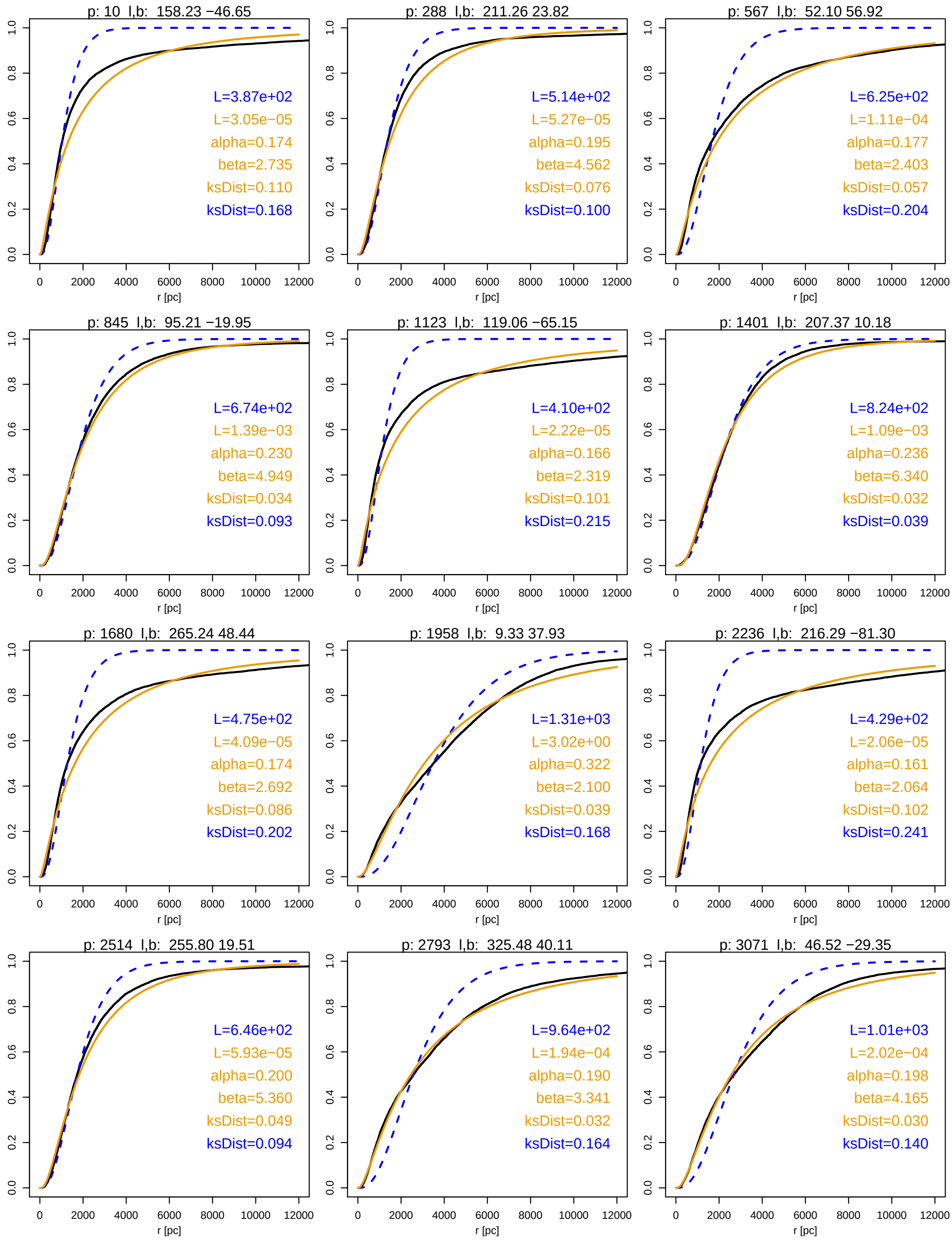

Figure 19: As Figure 18, but now shown as cumulative distributions.

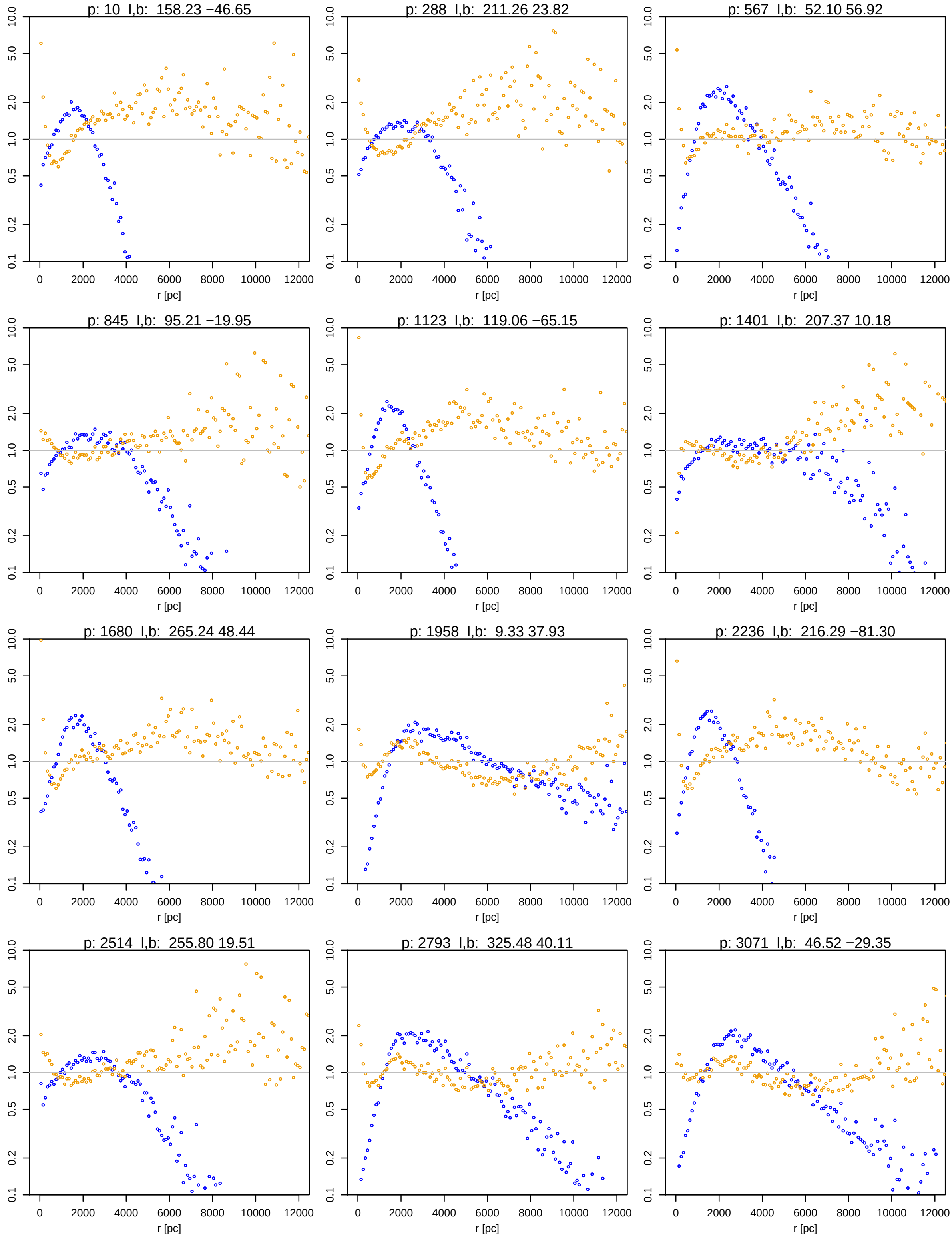

Figure 20: As Figure 18, but now showing the ratio of the GGD prior to the data (orange solid line) and the ratio of the EDSD prior to the data (blue dashed line). The vertical axis is logarithmic.

9 Distance inference given a prior on M_G

Given a prior $P(M_G)$ in addition to the prior $P(r)$, we would like to determine the posterior $P(r \mid \varpi, M_G, \ldots)$, where $\ldots$ represents other measurements or parameters on which the prior may depend. The absolute magnitude can be introduced into the inference using the relation

$$M_G + A_G \;=\; G - 5\log_{10} r + 5 \,. \tag{9}$$

If we know A_G and G, a prior on M_G constrains r. Let us assume $A_G = 0$. Alternatively we could assume we have an extinction-corrected G or that our prior is on $M_G + A_G$. We are therefore interested in determining the posterior $P(r \mid \varpi, G, \ldots)$. The paper IV distance prior can be considered as $P(r \mid l, b, \Phi)$, where Φ represents the galaxy model (including dust) and the Gaia selection function, and possibly also the colour of the star (taken here as noise free). This selection function tells us, first and foremost, the probability of observing a star of some apparent magnitude. Note, however, that in paper IV this selection function is applied solely to the model (and not the measured data) to influence the value of the length scale $L(l, b)$.

In what follows I will make the dependence on (l, b) implicit, but I will retain the explicit dependence on Φ.

From Bayes' theorem, the unnormalized posterior is

$$P^*(r \mid \varpi, G, \Phi) \;=\; P(\varpi \mid r, G)\, P(r \mid G, \Phi) \,. \tag{10}$$

The first term on the right side is the usual parallax likelihood, and is independent of Φ. Its dependence on G feeds into the model for σ_ϖ. Equivalently, we can assume σ_ϖ is a given, noise-free parameter on which this likelihood term depends, and then drop its dependence on G, i.e. it becomes $P(\varpi \mid r, \sigma_\varpi)$. The second term we can write as a marginalization over M_G, and then apply Bayes' theorem

$$\begin{aligned} P(r \mid G, \Phi) \;&=\; \int P(r, M_G \mid G, \Phi)\, \mathrm{d}M_G && (11)\\ &=\; \int \frac{1}{P(G \mid \Phi)}\, P(G \mid r, M_G)\, P(r, M_G \mid \Phi)\, \mathrm{d}M_G && (12)\\ &=\; \frac{P(r \mid \Phi)}{P(G \mid \Phi)} \int P(G \mid r, M_G)\, P(M_G \mid r, \Phi)\, \mathrm{d}M_G && (13) \end{aligned}$$

The first term in the integral is the likelihood for G (and is conditionally independent of Φ due to equation 9). In Gaia, G will almost always be much more precisely measured ($\lesssim 0.01$ mag) than the prior on M_G can be specified. In this case, the integral is non-zero only when equation 9 is satisfied. Hence the (unnormalized) posterior is

$$P^*(r \mid \varpi, G) \;=\; P(\varpi \mid r, \sigma_\varpi) P(r \mid \Phi) P(M_G = G - 5\log_{10} r + 5 \mid \Phi) \tag{14}$$

(the missing normalization constant, $1/P(\varpi, G \mid \Phi)$, is not usually required). This posterior is the familiar distance posterior of papers I–IV, but now multiplied by an additional prior over

M_G that is evaluated at a distance r and a measured apparent magnitude G. Note this prior's dependence on Φ: the measured apparent magnitude now has an influence, via the selection function, on the probability of the star having this particular M_G, and therefore it being at this particular distance r.

How we proceed from here depends on how we specify the prior on M_G. The computationally simplest approach is to adopt an explicit prior on M_G, perhaps provided directly by the user for a set of stars of known type. For example, if the prior were a Gaussian of mean $\widehat{M_G}$ and standard deviation σ_{M_G} (this being the information from Φ), then this prior would be written

$$P(M_G = G - 5\log_{10} r + 5 \mid \Phi) \;=\; \frac{1}{\sqrt{2\pi}\sigma_{M_G}} \exp\left(-\frac{(G - 5\log_{10} r + 5 - \widehat{M_G})^2}{2\sigma^2_{M_G}}\right) \;. \quad (15)$$

More generally, we would want to build a model for this prior that depends on other observational quantities, such as colour or position in the Galaxy. This should also accomodate the observational selection function. Most simply, the selection function can be written as $S(G)$, the probability of observing a source at magnitude G. Note that this is a probability, not a probability *density*. We can introduce such a selection function simply by multiplying the prior over M_G by $S(G)$. This is then unnormalized, but provided the normalization constant is independent of r, it is unimportant.[7] This selection function must of course be the same one as used to construct the other part of the prior, $P(r \mid \Phi)$.

We could construct the overall prior in a similar way as we did in paper IV:

1. Choose a 3D Galaxy model of stars with determined M_G, as well as a 3D extinction model and selection function (for example the planned Gaia DR3 mock catalogue).

2. Using the Galaxy model, identify all the stars along a narrow line-of-sight centered on (l, b).

3. Using the extinction model and the selection function, determine which of these stars are observed and what their (noise-free) apparent magnitudes (G) are.

4. Determine the parameters of the distance prior, $P(r \mid \Phi)$, using just the values of distances (note that this has been affected by the selection function). This is just the scalar L for the EDSD and EDD priors.

5. Determine the parameters of the absolute magnitude prior, $P(M_G \mid r, \Phi)$, using the values of the absolute magnitudes and, in general, the distances. However, if the parameters of this prior are functions of distance, then after applying the selection function, the normalization constant of this prior becomes a function of distance (as explained in the previous paragraph). It is desirable to avoid this by adopting a distance-independent prior over M_G.

[7] For given G, equation 15 is a Gaussian PDF in $\log r$. Consider $S(G)$ to be a top hat function. Multiplying this by the prior leaves a portion of a Gaussian, the area under the portion retained – the normalization constant – being a function of the parameters of the Gaussian, namely $\widehat{M_G}$ and σ_{M_G}, but not of r. This assumes, however, that neither $\widehat{M_G}$ nor σ_{M_G} are functions of r.

6. Repeat steps 2–5 for all lines-of-sight.

7. Possibly now fit a model to the parameters of the two priors to make them smooth functions of (l, b); this was done in paper IV using spherical harmonics.

This approach has implicitly assumed someting about stellar physics and stellar populations, because the 3D Galaxy model has provided (a) self-consisent measures of G, M_G, A_G, and r for individual stars, (b) a reasonably realistic distribution of types of stars in the HRD. Indeed, if we set $\Phi = \{G, \mathrm{BP}-\mathrm{RP}\}$ then this is what we often call the HRD prior. This is essentially the same as what we did for the Galaxy model in paper III (section 4), although there we viewed this as likelihood on the observed magnitude and colour (equation 17 of that paper) rather than a prior on M_G. The only real difference is whether we take into account the uncertainties in the magnitude (and more generally also the colours), in which case the integral in equation 13 is no longer non-zero only when equation 9 is satisfied, meaning we would have to actually do the integral (although it could be approximated).

I now demonstrate this method of distance inference using a set of red clump stars in GDR2 that were identified by Ting et al. (2018) (this paper was accepted to the journal before GDR2 appeared). For simplicity I assume a separable prior and no explicit selection function $S(G)$ (although one was used in the building the length scale model, so in principle there is an inconsistency here). I consider just the APOGEE `RC_Pristine` sources listed in their Table 2. This contains 36 908 sources, of which 36 788 have pre-computed crossmatches in GDR2 (via 2MASS). I use the same distance prior as in paper IV and adopt a simple (but perhaps not very realistic) Gaussian prior on absolute magnitude (i.e. equation 15) with $(\widehat{M_G}, \sigma^2_{M_G}) = (0.5, 0.5)$ mag. Examples of some posteriors are shown in Figure 21. I sampled the posterior using the Metropolis algorithm for 2000 iterations (after 100 burn-in steps) initialized at $1/\varpi$ with a step size equal to $\frac{1}{\varpi}\min(\sigma_{M_G}, \frac{\sigma_\varpi}{\varpi})$. From the resulting samples I computed the median, as well as the 0.159 and 0.841 quantiles which together form the 68.2% equal-tailed confidence interval (ETI).

We see from Figure 21 how the posterior is a compromise between the naive parallax-only and naive absolute-magnitude-only distance estimates. The length scale of the prior seems to be playing little role (the mode of the prior is at $2L$), even for the largest fpu plotted (second row, middle column). A comparison of the inferred distances (using the median) to those from paper IV for 180 of these red clump stars is shown in Figure 22 (all the parallaxes happen to be positive). The absolute magnitude prior is producing larger distance estimates for more distant stars.

Figure 23 compares the width of the posteriors (the size of the 68% ETI) using the absolute magnitude prior in addition to the EDSD prior to the width of the posteriors using just the EDSD prior. The median of the ratios shown is 1.02 (68% ETI is 0.91–1.18). This suggests that for these stars the distance estimates are not actually more *precise* when using this absolute magnitude prior for these stars, but they may still be more *accurate* (which we cannot reliably test here, as we do not know the true distances).

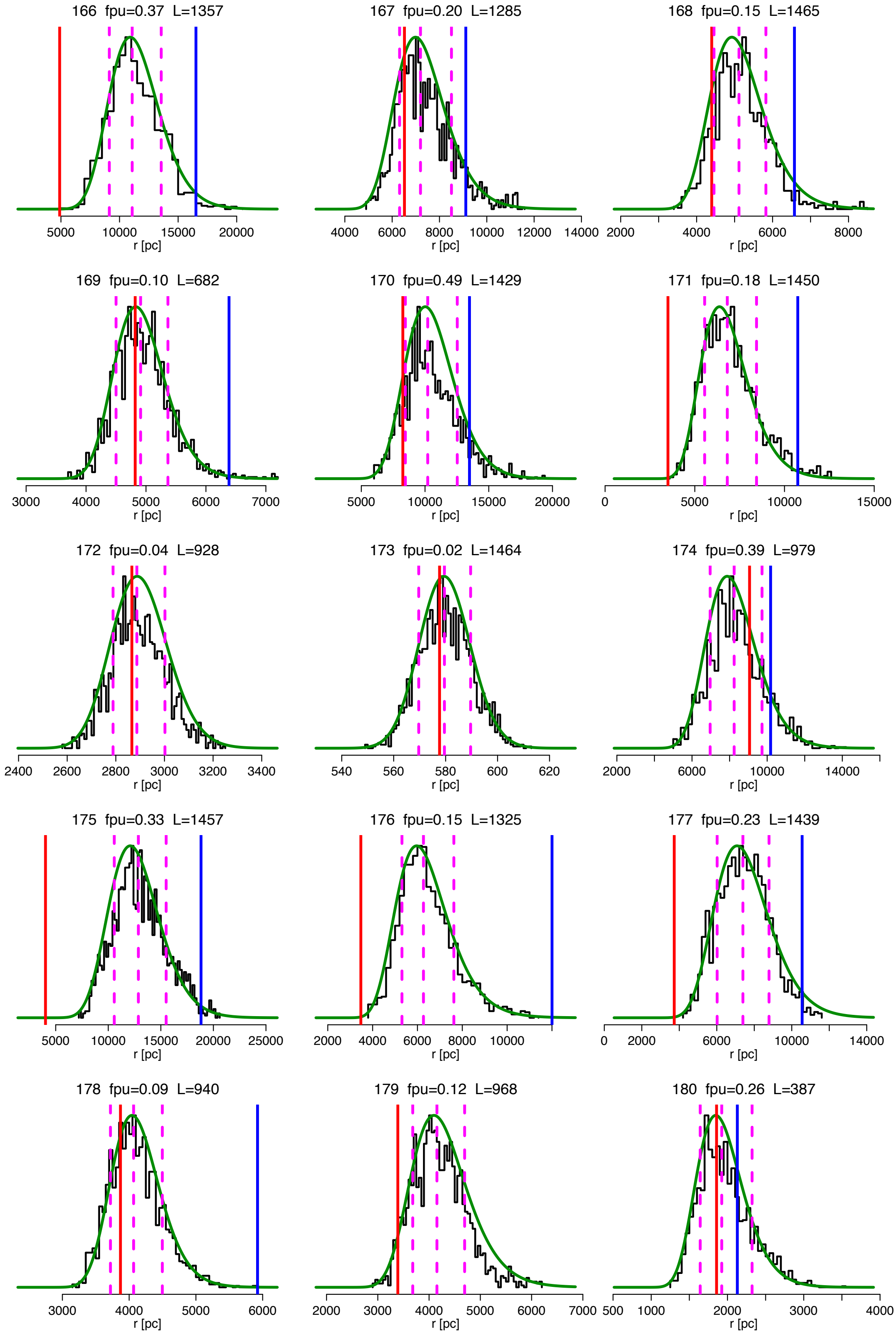

Figure 21: Examples of posteriors computed using the EDSD distance prior and absolute magnitude prior $(\widehat{M_G}, \sigma^2_{M_G}) = (0.5, 0.5)$ mag. Green is the posterior, the black histogram is the MCMC sampling thereof (scaled to have the same maximum). The vertical magenta lines show the 0.159, 0.5, and 0.841 quantiles. The vertical red line is $1/\varpi$. The vertical blue line is $10^{(G-M_G+5)/5}$, the absolute-magnitude-only distance estimate.

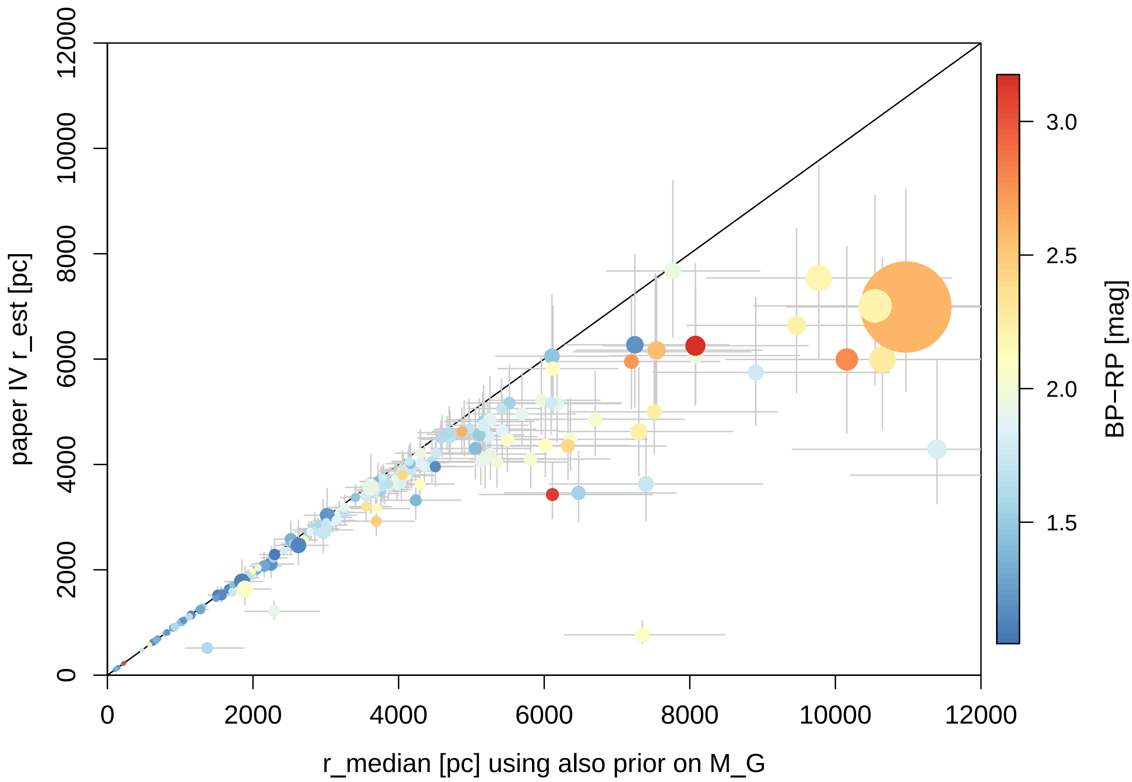

Figure 22: Comparison of distance estimates for red clump stars from paper IV (vertical axis) and inferred here using an additional absolute magnitude prior (horizontal axis). The error bars are the (asymmetric) 68% confidence interval (the HDI for the vertical axis and the ETI for the horizontal axis). The colour of each point denotes its BP-RP colour (for information; not used in the inference). The area of each point is proportional to the fpu (with the largest equal to 8.8).

10 Conclusions

The main findings are as follows.

1. The median of the distance posterior corresponds to the same distance as the median of the log distance posterior (using the same prior on distance). Thus a single distance estimate can be provided which can be used also for the distance modulus (and any monotonic transformation of the distance). When providing the median we should provide equal-tailed confidence intervals instead of the highest density interval. For GDR2, using the median rather than the mode has no impact on the estimated distance in the limit of negligible parallax uncertainties. In the limit of very large parallax uncertainties, the distance is increased by between 20% and 40% in most cases.

2. The GGD prior, with three parameters, is more flexible than the EDSD prior (which is a special case of it). Unsurprisingly, the GGD prior fits the data in the GDR2 mock catalogue better. The price we pay is the loss of a clear physical interpretation of its

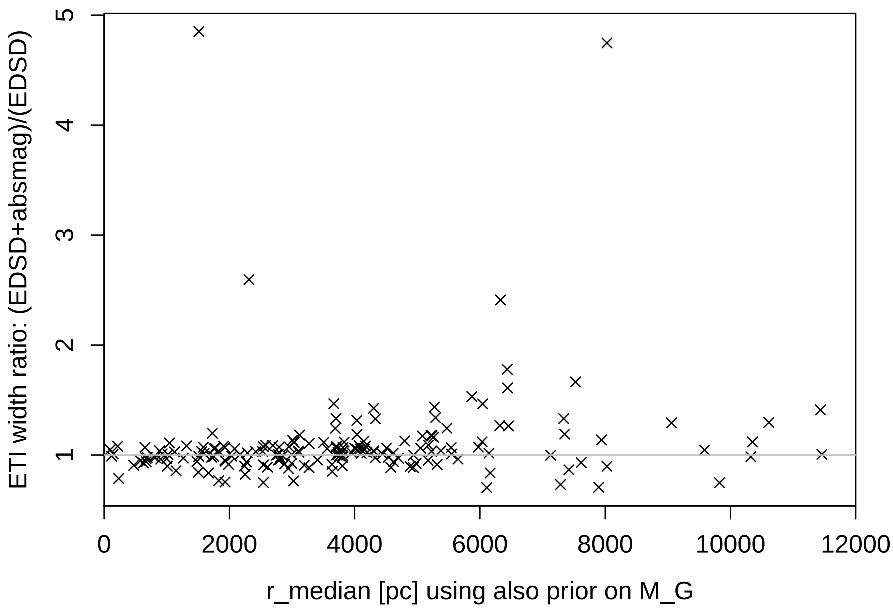

Figure 23: Ratio of the width of the 68% ETI when using the absolute magnitude prior together with the EDSD prior to that using just the EDSD prior. This is for the same stars as shown in Figure 22.

parameters. The EDD prior, which is a simplified version of the EDSD prior, is viable, and has the advantage that the posterior is unimodal.

3. Computing quantiles by MCMC may not be time prohibitive. With the EDSD prior and 2000 samples, it would take around 600 CPU-days for 1.33 billion stars. This is only eight times longer than required for the paper IV computations. A more detailed analysis of the quality of the MCMC sampling is required, however, to optimize the number of samples.

4. Geometric distances can be easily and sensibly improved by placing a prior on M_G and using the measured G (even if assumed noise-free). Such photogeometric distances are not necessarily more precise (depending on the prior used), but they should be more accurate in the limit of poor parallaxes.

5. There is a dependence of distance on G and $\mathrm{BP}-\mathrm{RP}$ that could be exploited in a more sophisticated prior than the (l, b)-only dependence used in paper IV. This is being explored.

11 Recommendations

1. We should provide a catalogue of geometric distances as in GDR2, either using the EDSD prior or with the GGD prior. The deciding the factor between the priors is the preference of interpretability and simplicity of the EDSD prior or the superior fit of the GGD prior. As the GGD prior involves three interdependent parameters, it seems that we could not

smooth these parameters over the sky – as we did for the single parameter of the EDSD prior in paper IV – to get get a sensible variation of the GGD prior over the sky. Instead we would have to use a fixed prior for a finite area (e.g. HEALpixel) which then jumps discontinuously to its neighbouring pixels.

2. When fitting a single length scale for the distance prior using a mock Gaia catalogue, then although it is prudent to use the median (rather than the mean) to avoid the negative impact of outliers, we should equate this to the theoretical median of the prior. This is easily done for the EDSD prior, as there is a tight empirical linear relationship between the median and mean, such that $L = 0.374\, r_{\rm median}$ (see section 6).

3. We should summarize the posterior using quantiles instead of the mode+HDI. At least the median and 16th and 84th (approximately) quantiles – that provide a "1sigma"-like ETI – should be provided. We could consider providing more quantiles, such as 1st/99th, 5/95th, and/or 2sigma, to better characterize the posterior. Extreme ones (like the 99.9th) would be too noisy (given the limited number of samples we have time to draw), so should not be provided. Adding more quantiles makes the catalogue a lot bigger, however. The GDR2 distance catalogue – with five columns and 1.33 billion rows – is about 130 GB, ASCII uncompressed, i.e. 26 GB per column. Nominally we would use an MCMC routine to sample the posterior, but some clever sampling on a grid may allow fewer evaluations. As we provide quantiles and not the mode, the unimodal benefit of the EDD prior is obsolete (it confers no other advantages).

4. We should provide a second set of distances that also take into account G, and probably $\rm BP-RP$ too. There are a least three ways of doing this: (1) we consider G and $\rm BP-RP$ to be error-free and simply extend the prior from $L(l, b)$ to $L(l, b, G, \rm BP-RP)$; (2) we introduce a likelihood in G that depends on the unknown M_G, upon which we place a (colour-dependent) prior and marginalize (equation 10 to 13), perhaps conditional also on the colour (taken to be error-free); (3) as (2), but also with a likelihood on $\rm BP-RP$ (as done in paper III section 4). However done, the distance prior should be the same one as in the geometric catalogue, and use the same posterior summary statistics. Work I will report on elsewhere shows that option (2) works well for GDR2. Given the small uncertainties in $\rm BP-RP$ compared to those in the M_G prior, option (3) probably brings little improvement over this.

5. Given the range of use cases of distances, we should provide, in addition to a catalogue, a software tool that can compute distance (and log distance) for a few different geometric priors (e.g. GGD, EDSD, EDD) with user-specified parameters, and perhaps also choice of estimator (mean, median, mode) and confidence intervals (standard deviation, ETI, HDI). This tool should include the option to specify an explicit prior on M_G (section 9).

12 References

Astraatmadja T., Bailer-Jones C.A.L., 2016a, Estimating distances from parallaxes II. Performance of Bayesian distance estimators on a Gaia-like catalogue, *Astrophysical Journal*, 832, 137 (paper II)

Astraatmadja T., Bailer-Jones C.A.L., 2016b, Estimating distances from parallaxes III. Distances of two million stars in the Gaia DR1 catalogue, *Astrophysical Journal*, 833, 119 (paper III)

Bailer-Jones C.A.L., 2015, Estimating distances from parallaxes, *Publications of the Astronomical Society of the Pacific*, 127, 994 (paper I)

Bailer-Jones C.A.L., 2017, *Proposal for providing distance estimates in GDR2*, GAIA-C8-TN-MPIA-CBJ-080

Bailer-Jones C.A.L., 2018, *Estimating stellar distance, extinction, and atmospheric parameters from a spectrum and parallax*, GAIA-C8-TN-MPIA-CBJ-087

Bailer-Jones C.A.L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, Estimating distances from parallaxes IV: Distances to 1.33 billion stars in Gaia data release 2, *Astronomical Journal*, 158, 58 (paper IV)

Rybizki J., Demleitner M., Fouesneau M., Bailer-Jones C.A.L., Rix H.-W., Andrae R., 2018, A Gaia DR2 Mock Stellar Catalog, *Publications of the Astronomical Society of the Pacific*, 130, 074101

Ting Y., Hawkins K., Rix H.-W., 2018, A large and pristine sample of standard candles across the Milky Way: 100,000 red clump stars with 3% contamination, *Astrophysical Journal Letters*, 858, 7