Re-normalising the astrometric chi-square in *Gaia* DR2

prepared by: L. Lindegren
affiliation : Lund Observatory
approved by:
reference: GAIA-C3-TN-LU-LL-124-01
issue: 1
revision: 0
date: 2018-08-16
status: Issued

Abstract

The second release of *Gaia* data (*Gaia* DR2) contains several statistical indicators that can be used to assess the quality and reliability of the astrometric data. One such indicator is the astrometric chi-square, or equivalently the unit weight error (UWE) being the square root of the reduced chi-square. The usefulness of the UWE is however severely hampered by its distribution having strong dependences on magnitude and colour. These dependences can be eliminated by a re-normalisation process, using tables provided with this TN. The re-normalised UWE (or RUWE) is a more reliable and informative goodness-of-fit statistic than for example the astrometric excess noise.

Document History

Issue	Revision	Date	Author	Comment
1	0	2018-08-20	LL	Expanded Sect 5 and added 6, revised Abstract, etc.
D	0	2018-07-10	LL	Creation and first draft.

Contents

1 Introduction

The second data release of the *Gaia* mission (Gaia Collaboration et al. 2016, 2018) contains astrometric five-parameter solutions for more than 1.3 billion sources, data which are publicly available through the *Gaia* Archive.[1] Although these solutions have passed a number of checks before publication, their quality varies enormously depending on many factors such as the magnitude of the source, the number of observations per source, the source environment (e.g. if it is in a crowded area), and the geometric properties of the source (e.g. if it is a point source, binary, or extended object). A question very often raised by users of the *Gaia* astrometry is how to distinguish between 'good' and 'bad' solutions. The *Gaia* Archive contains several statistics that can be used for this purpose, but their interpretation is not always straightforward and some are more useful than others.

The purpose of this note is to examine one of the more useful statistics implicit in the archive data, namely the 'unit weight error' (UWE) u, and to provide a recipe for its use. The UWE is not explicitly given in the *Gaia* Archive but can easily be calculated as described in Sect. 3.

Theoretically, the UWE is expected to be close to 1.0 for well-behaved solutions of single stars, but in practice it is often found to be much larger even for solutions that seem to be perfectly good. This happens in particular for sources of extreme colours, or that are very bright. To be practically useful the UWE needs to be scaled by a factor depending on the magnitude and colour of the source. This is derived in Sect. 4. The resulting re-normalised UWE, or RUWE, behaves much more uniformly, so that for example a single threshold can be used, at all magnitudes and colours, to distinguish between 'good' and 'bad' solutions.

A comment should be made here about the relation between the RUWE and the formal uncertainties ('standard errors') of the astrometric parameters. The two kinds of statistic provide orthogonal information in the sense that small formal uncertainties are not necessarily accompanied by a good RUWE, and vice versa. However, they are not independent either. The formal uncertainties reflect the internal consistency of the measurements and consequently encapsulate all sorts of errors that create a mismatch to the five-parameter model. This includes for example the increased scatter for astrometric binaries. The RUWE should therefore *never* be applied as a 'correction factor' on the formal uncertainties: it is, so to speak, already taken into account. Although the true (external) errors are generally underestimated by the formal uncertainties, the required correction can only be derived from comparisons with independent (external) data, which is not the subject of this note.

[1] `https://archives.esac.esa.int/gaia`

2 Astrometric quality indicators in *Gaia* DR2 and their uses

Quality indicators could have many different uses. Some possible scenarios are:

1. We want sources with the most *precise* astrometric data. In this case the selection should simply be made based on the formal uncertainties of the relevant parameters (e.g. `parallax_error`).
2. We want sources with the most *reliable* astrometric data. One interpretation of reliability in this context could be that the formal uncertainties should reasonably reflect the actual error distribution; in particular, we may want to avoid data that are wrong by many standard deviations. The most relevant statistic for this selection is probably the number of visibility periods (`visibility_periods_used`): a higher number of visibility periods means that the solution is less sensitive to the occasional bad measurement.
3. We want sources for which the observations are *consistent* with the astrometric five-parameter model, suggesting a well-defined centre of light with uniform motion over the observation interval, and negligible disturbance from neighbouring sources.
4. We want sources whose observations are *inconsistent* with the astrometric five-parameter model, where the discrepancy could for example be caused by binarity.

The UWE discussed in this note is mainly relevant in the last two scenarios, where it provides an astrometric goodness-of-fit whose interpretation is comparatively simple. However, it is also relevant as a secondary criterion in the first two scenarios, since a large UWE could mean that the observations are disturbed (e.g. by the proximity of another source), which could make the data less precise or less reliable.

An alternative goodness-of-fit statistic available in *Gaia* DR2 is the excess source noise, `astrometric_excess_noise`, with its significance `astrometric_excess_noise_sig`. In theory, the excess source noise encodes similar information as the UWE, but expressed as an angle (in mas) rather than in relation to the expected noise. This makes it less straightforward to interpret than the UWE: a large excess noise does not necessarily mean that there are strong deviations from the five-parameter model, namely if the significance is low. The excess noise is also less informative in the many cases when it is zero or insignificant.

A third potential quality indicator is the fraction of along-scan outliers, calculated as astrometric_n_bad_obs_al/astrometric_n_obs_al. However, because the outlier rejection algorithm was designed to identify a small number of strongly deviating observations, this statistic is relatively insensitive to model mismatches, which typically affect most of the observations of a source. The outlier fraction is also uninformative when the number of outliers is zero.

Finally, there is the 'gaussianised' goodness-of-fit, $F2 = \texttt{astrometric\_gof\_al}$, which applies the cube-root transformation (Wilson & Hilferty 1931)

$$F2 = \left(\frac{9\nu}{2}\right)^{1/2}\left[\left(\frac{\chi^2}{\nu}\right)^{1/3} + \frac{2}{9\nu} - 1\right] \tag{1}$$

such that $F2$ is approximately unit normal $N(0,1)$ if $\chi^2 = \texttt{astrometric\_chi2\_al}$ follows the chi-square distribution with $\nu = N - 5$ degrees of freedom, where $N =$ astrometric_n_good_obs_al is the number of observations. Ideally, this distribution is expected for well-behaved solutions of single stars, but in practice that is often not the case and the gaussianised goodness-of-fit is then rather pointless.

In conclusion, for the various reasons outlined above, we advise against using these statistics for the general discrimination of 'good' and 'bad' solutions, and instead advocate the use of the simpler UWE, or even better the re-normalised version of it (RUWE) explained below.

3 The unit weight error in *Gaia* DR2

For a five-parameter solution in *Gaia* DR2 ($\texttt{astrometric\_param\_solved} = 31$) we refer to the statistic

$$u = \sqrt{\frac{\chi^2}{N-5}} = \sqrt{\frac{\texttt{astrometric\_chi2\_al}}{\texttt{astrometric\_n\_good\_obs\_al} - 5}} \tag{2}$$

as the 'unit weight error' (UWE) of the source. There is no generally accepted name for this statistic; sometimes it is called the standard error of unit weight, but 'unit weight error' seems to have become relatively standard at least in the astrometric community. Its square u^2 is the reduced chi-square statistic, or the unit weight variance (UWV). In Eq. (2) N is the number of good along-scan observations of the source (i.e. not counting outliers), and χ^2 is the corresponding sum of the squares of the along-scan residuals R_l, divided by their standard uncertainties:

$$\texttt{astrometric\_chi2\_al} = \chi^2 = \sum_{l=1}^{N}\left(\frac{R_l}{\tilde{\sigma}_l}\right)^2 . \tag{3}$$

Naively we expect well-behaved solutions to have $\langle\chi^2\rangle \simeq N - 5$, and consequently $\langle u\rangle \simeq 1.0$. Even if a certain fraction of the solutions are bad (e.g. because of duplicity), we expect the distribution of u to peak roughly at 1.0, and the median u should be only slightly greater than unity. Plotting the distribution of u for a few different magnitude ranges (Fig. 1) shows that this is approximately true only for faint sources. Although the overall shape of the distribution is roughly as expected, it is shifted towards higher values for bright sources ($G \lesssim 13$) and to smaller values for intermediate magnitudes.

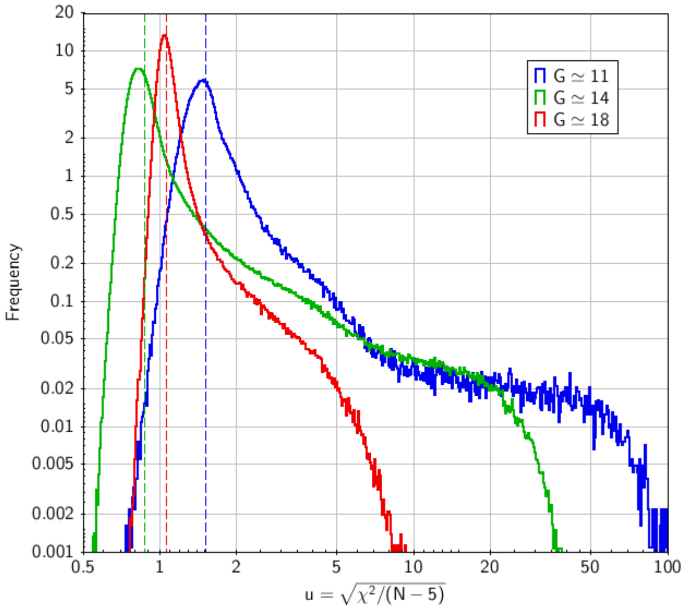

FIGURE 1: Histogram of the UWE (u) for five-parameter solutions of sources in *Gaia* DR2 with G magnitude in the ranges 10.8–11.2, 13.9–14.1, and 17.995–18.005. The dashed vertical lines show the medians of the distributions.

From Fig. 1 it appears that a more reasonable distribution of the UWE could result from simple scaling by an appropriate factor depending on magnitude. An appreciation of the magnitude effect is obtained from Fig. 2, where the individual and median UWE values are plotted versus G. The strong variation around $G = 13$ is caused by the predominant use of two-dimensional CCD windows for observations brighter than this limit, and one-dimensional windows for the fainter observations. Other features are produced by the different CCD gates (for $G \simeq 10$ to 12) and saturation ($G \lesssim 6$). Clearly the scaling factor is a rather complicated function of G.

There is however a very strong dependence on the colour index $C = G_{\rm BP} - G_{\rm RP}$. This is illustrated by Fig. 3, showing the median UWE as a function of both G and C. It is now seen that the general variation with G suggested by the median curve in Fig. 2 is only valid for intermediate colours, roughly in the interval $C \simeq 1.0$ to 2.0; for bluer and redder sources the median UWE quickly rises to values that could approach 10 for very red sources. This shows that the astrometric calibration of the *Gaia* instrument did not completely remove the chromatic effects.[2] The scaling factor must therefore depend on (at least) both magnitude and colour.

Should the UWE be corrected for other factors besides the magnitude and colour? The tentative answer to this is 'no'. The pragmatic reason is that there are not enough sources to make a reliable statistical analysis of the UWE if they should be subdivided by a third factor after G and C. But there is also a good theoretical argument for this answer. The scaling of the UWE is intended to compensate for deficiencies in the calibration model. It should then not include dependencies that are not already used in the calibration model, or that could potentially have been used. Examples of such dependencies are the time of the observation (because the instrument evolves), the CCD index of the observation (because each CCD behaves differently), and the field of view (different parts of the optics are used for the two fields of view). But these are all properties of the individual observation, and cannot sensibly be described as a property of the source.

Will the UWE then be independent of factors such as the number of observations and the position on the sphere? Not necessarily, but we may not want to take them into account even if there is such a dependence. Figure 4 shows that there is very little dependence on the number of observations. Although this plot is for sources around $G = 14$ mag and intermediate colours, it is fairly representative for any well-populated region of the colour-magnitude space. Figure 5 shows the variation of the median UWE across the celestial sphere for the same sample of sources as in Fig. 4. Although some features are clearly related to the scanning law (the bluish or yellow/red arcs seen most easily in the Galactic projection), most of the variations are rather related to Galactic features such as the density of sources. If there is a variation depending on the environment, this is something that should not be corrected as it will help to single out perturbed sources. In any case the amplitude of the variations seen in these maps is moderate ($\pm 15\%$) compared with the colour dependence in Fig. 3. Thus we conclude that no other factor need to be taken into account in addition to the magnitude and colour.

[2]This may seem surprising given that the astrometric calibration model included chromatic terms depending on the effective wavenumber $\nu_{\rm eff}$; see Sect. 3.3 in Lindegren et al. (2018). The typical size of these chromatic terms was of the order of 2 to 10 mas μm (see Fig. 3.19 in the *Gaia* DR2 online documentation, Sect. 3.5.5), corresponding to colour-dependent shifts of 0.8 to 4 mas for a change in effective wavenumber from an average $\nu_{\rm eff} = 1.6$ to 1.2 μm^{-1} for a very red star (see Fig. 1 in Lindegren et al. (2018) for the relation between effective wavenumber and colour index). By contrast, if a bright ($G \simeq 10$), red ($C \simeq 5$) source obtains a typical UWE of about 5 according to Fig. 3, this only corresponds to an RMS residual of about 0.4 mas, since the along-scan uncertainty for such a source is $\tilde{\sigma}_l \simeq 0.08$ mas. This example shows that most of the chromatic shifts must indeed have been removed by the chromaticity calibration in the astrometric solution, and that one only needs a remaining smaller part of the chromaticity to explain the increased UWE seen for blue and red stars in Fig. 3.

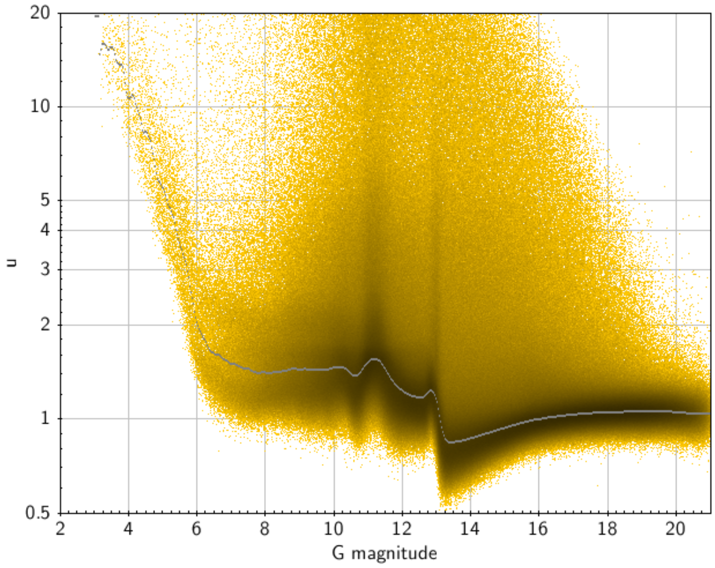

FIGURE 2: UWE (u) for a subset of five-parameter solutions in *Gaia* DR2 as a function of the G magnitude. The yellow dots are for the individual solutions; the grey curve is the (slightly smoothed) median. The subset include all sources for $G \leq 11$ mag, and an exponentially decreasing random fraction of the fainter sources.

4 The reference UWE (u_0) for well-behaved solutions

Based on the discussion in Sect. 3 we assume that the re-normalisation of the UWE can be achieved by the a simple scaling,

$$u_{\text{norm}} = u/u_0(G, C)\,, \tag{4}$$

where $u_0(G, C)$ is a reference value that remains to be determined. We refer to u_{norm} as the re-normalised unit weight error, or RUWE. Naturally, the re-normalised reduced chi-square statistic is simply u^2_{norm}, and the re-normalised chi-square statistic is $(N-5)u^2_{\text{norm}}$. However, it cannot be assumed that these quantities follow the standard statistical distributions for well-behaved solutions (see Sect. 5).

Clearly $u_0(G, C)$ should be determined in such a way that $\langle u_{\text{norm}}\rangle \simeq 1.0$ for well-behaved solutions of single stars. Effectively, this means that $u_0(G, C)$ should equal the mean (or median) UWE for well-behaved solutions of single stars. The problem is that there is no practical way to construct a sample of well-behaved solutions of single stars at any magnitude or colour, let alone as a function of magnitude and colour. Instead, for any (small) range of magnitude and colour, we must use the full sample of solutions, well-behaved or not, and rely of statistical arguments to estimate the average u_0 in the magnitude–colour bin.

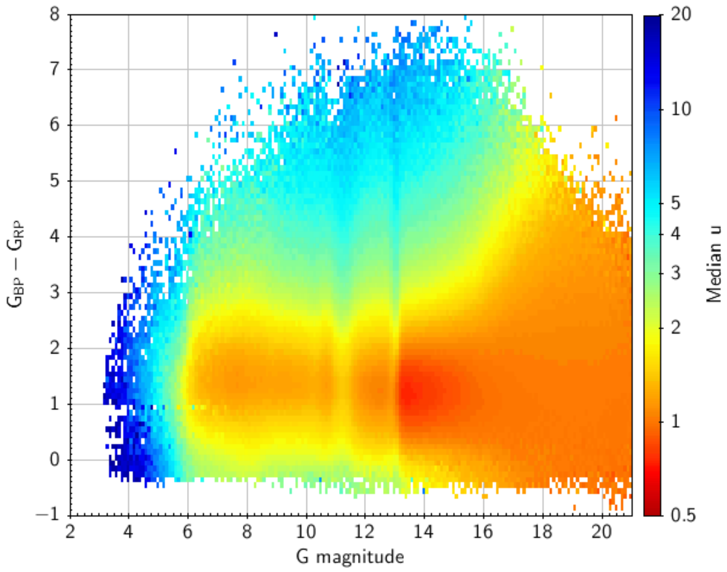

FIGURE 3: Median UWE (u) as a function of magnitude (G) and colour ($C = G_{\rm BP} - G_{\rm RP}$) for a subset of the sources in Fig. 2 with good colours. The bin size is 0.1 mag in both coordinates. Only sources with 'good' colours were used for this diagram (and subsequent analysis), i.e. where the flux excess factor satisfies Eq. (C.2) in Lindegren et al. (2018).

It is necessary that $u_0(G, C)$ is a reasonably smooth function in both arguments, lest artefacts in the distribution of the RUWE are introduced by the re-normalisation. Moreover, $u_0(G, C)$ should be available even in sparsely populated areas of the magnitude–colour space. These requirements mean that semi-analytical functions such as splines must be used to fit, smooth, and interpolate the values of u_0 obtained per magnitude–colour bin. We leave this somewhat messy procedure for later and focus now on how to estimate u_0 for a given magnitude–colour bin containing a reasonable number of sources.

One obvious thought is to use the median UWE (in the bin) as an estimate of u_0. However, we know that any sample contains a significant fraction of binaries, some of which will have an increased UWE, and also some solutions that are 'bad' for various other reasons. This means that the median u necessarily overestimates u_0. Differently put, there exists for every bin a certain percentage P, such that u_0 equals the Pth percentile of u, where $P < 50\%$. Is there some way to estimate P?

The raw distributions of u in Fig. 1 already provide a hint of what can be done. They show a relatively well-defined peak in the distributions of u slightly to the left of the median. Is the peak (mode) a good estimate of u_0? Theoretically, the distribution of u for the subsample of well-behaved single-star solutions is expected to peak very close

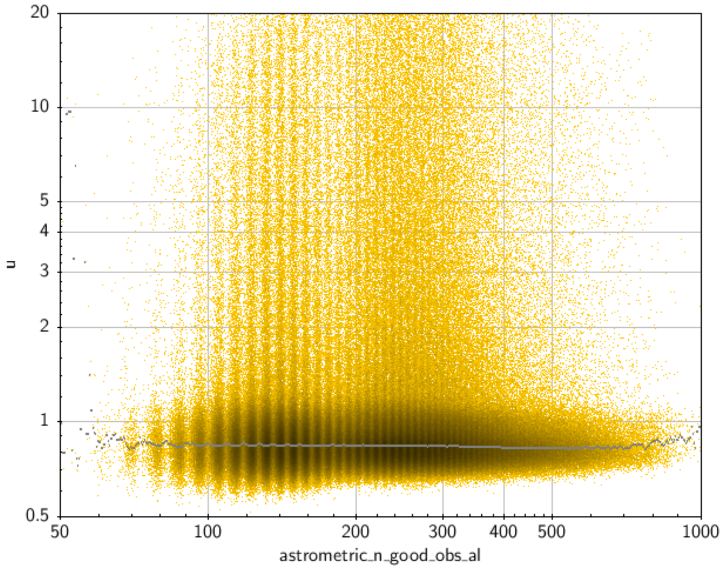

FIGURE 4: The UWE (u) as a function of the number of observations for five-parameter solutions with G magnitude in the range 13.9–14.1 and colour index C in the range 1.0–2.0. The yellow dots are for the individual solutions; the grey curve is the (slightly smoothed) median.

to the mean (or median) value. Adding on top of this the (unknown) distribution of not-well-behaved solutions, which is skewed towards larger values of u but contains some around unity as well, will shift the mode slightly towards larger values. However, this shift is much smaller than the difference between the mode and the median of the full sample. (This statement can be verified by numerical experiments.) Thus we conclude that the mode is a much better estimate of u_0 than the median, but still could be a slight overestimation of u_0. The mode is however problematic for two reasons. The first (but not the worst!) problem is that the location of the mode is not invariant under non-linear transformations. For example, the mode of u, u^2, and $\log u$ are all slightly different. Fortunately the differences are small in practice, and there is anyway a simple theoretical solution. It is a well-known property of the chi-square distribution that the cube root of χ^2 is Gaussian, to a considerable degree of approximation (Wilson & Hilferty 1931), and therefore approximately symmetrical with respect to the mode. Hence it makes sense to use the specific transformation $u^{2/3}$ to locate the mode.

The second problem is a much more serious one: to determine precisely the location of the mode requires a large sample. In practice it cannot be reliably determined for the small samples obtained with a reasonable binning in magnitude and colour, except possibly in the most densely populated areas of the magnitude–colour space. The approach taken here is pragmatic: use a fixed percentile P as a proxy for the mode, and estimate this P by means of some well-populated regions of the magnitude–colour

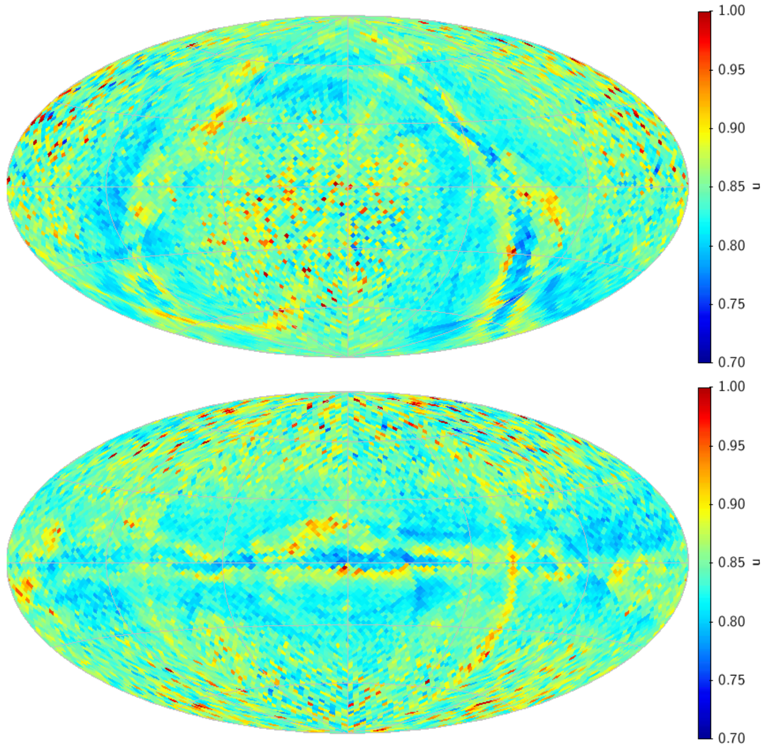

FIGURE 5: *Top:* Map of the median UWE (u) for five-parameter solutions with G magnitude in the range 13.9–14.1 and colour index C in the range 1.0–2.0. Median values of u are shown in pixels of about 3.36 deg^2. This full-sky map uses a Hammer–Aitoff projection in equatorial (ICRS) coordinates with $\alpha = \delta = 0$ at the centre, north up, and α increasing from right to left. *Bottom:* Same map in a Galactic projection with $l = b = 0$ at the centre, Galactic north up, and l increasing from right to left.

space. The resulting estimates of P range from 37% to 45%. For the remainder of this note we adopt $P = 41\%$. Thus u_0 is simply estimated as the 41st percentile of u in the relevant magnitude–colour bin. This was done for $3.6 \leq G \leq 21.0$ and the range of colours represented by the data at each magnitude.

Figure 6 shows the resulting estimates of u_0, plotted against the bin centres in G and C. To construct a continuous function $u_0(G, C)$ the following steps were taken. First, for each magnitude interval (i.e. a fixed G representing the bin centre along the magnitude axis), the variation $f(C) = u_0(G, C)^2$ was fitted by a quadratic spline, having knots at $C = 0.2$, 0.6(0.3)3.0, 3.5 and using max$(0, n - 2)$ as the statistical weights, where n is the number of ’sources in the bin. To stabilise the fit, the spline was constrained to be convex everywhere, $\partial^2 f/\partial C^2 \geq 0$, which seems to be consistent with the data at all magnitudes. A few examples of the spline fits are shown in Fig. 7. Secondly, the spline coefficients thus obtained were smoothed as functions of G and interpolated to a regular grid of magnitudes, $G = 3.6(0.01)21.0$. This was done using a quadratic

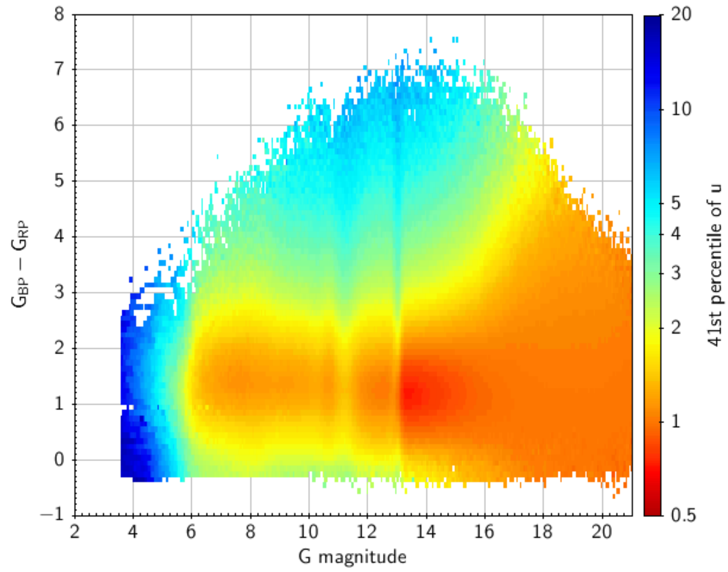

FIGURE 6: The reference UWE (u_0) estimated as the 41st percentile of u in bins of G and $C = G_{\rm BP} - G_{\rm RP}$. The sample is the same as in Fig. 3, but the binning is slightly difference (bin size in G increasing for $G < 6$), and only bins with at least three sources are displayed.

LOESS smoothing with a span of ± 0.11 mag in G. Using the smoothed spline coefficients then made it possible to compute a continuous function $u_0(G, C)$ for any C over the whole rage of magnitudes $3.6 \leq G \leq 21.0$. Although this gave acceptable results for a restricted range of colours (roughly $-0.5 < C < 5$ it resulted in strong oscillations versus G for more extreme colours due to the extrapolated splines. To eliminate this drawback, the cuts $u_0(G, -2)$ and $u_0(G, 10)$ were constrained to be smooth functions of G in a second fit of the splines. The resulting two-dimensional fit is shown in Fig. 8. For $C \gtrsim 5$ there remain small oscillations versus G, not supported by the data, but as they occur mainly in areas almost devoid of sources, they are of no practical importance.

Figure 9 shows the 41st percentile of the RUWE ($u_{\rm norm}$) obtained by applying Eq. (4) to the sample in Fig. 3. The process has successfully removed nearly all systematic variations with G and C. Remaining variations are of the order of a few per cent for $G \gtrsim 6$ and $0 \lesssim C \lesssim 4$ and a bit larger for more extreme colours.

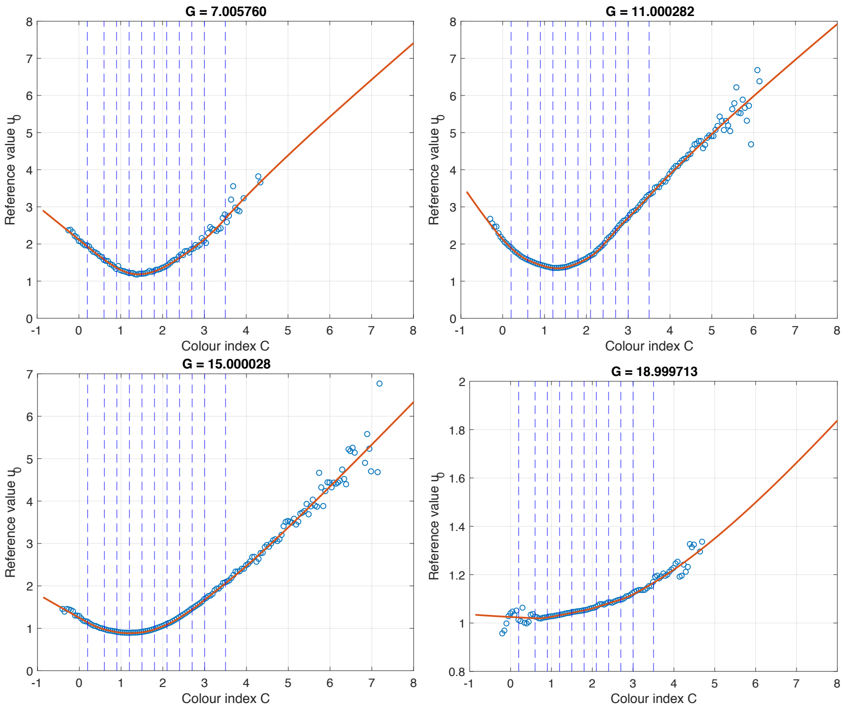

FIGURE 7: Examples of the fits of convex quadratic splines to u_0 versus $C = G_{\rm BP} - G_{\rm RP}$. The circles are the 41st percentiles of u in magnitude-colour bins; the red curve is the fitted spline. The dashed vertical lines show the positions of the knots. The four plots are for the magnitude intervals $G = 7.00 \pm 0.05$, 11.00 ± 0.05, 15.00 ± 0.05, and 19.00 ± 0.05 mag, using approximately 2k, 112k, 69k, and 56k sources, respectively.

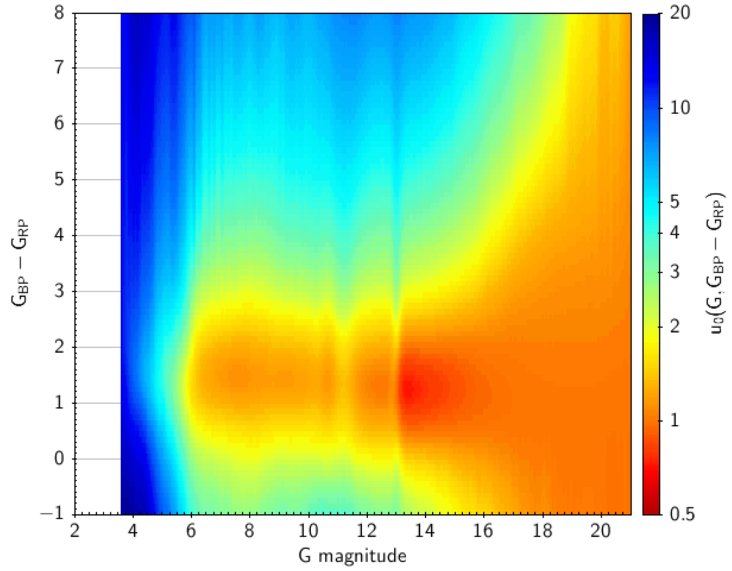

FIGURE 8: The fitted function $u_0(G, C)$ calculated for the whole range of G and for $-1 \leq C \leq 8$.

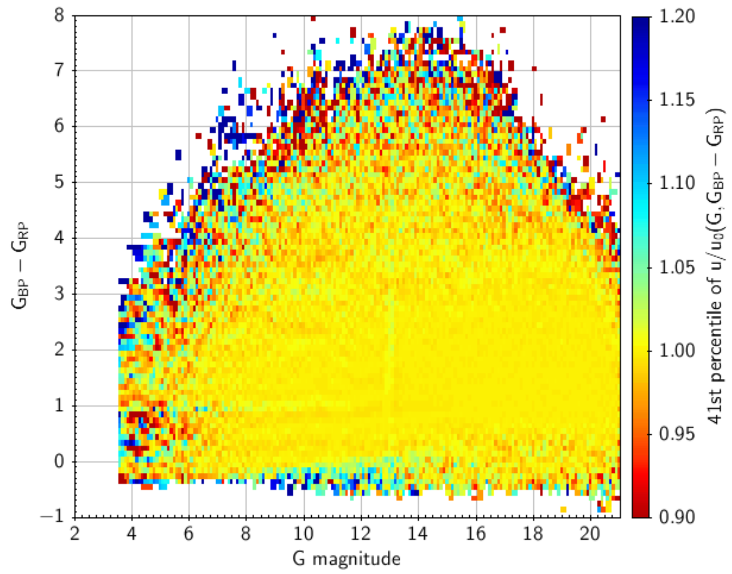

FIGURE 9: The 41st percentile of $u_{\rm norm}$ for the sample in Fig. 3.

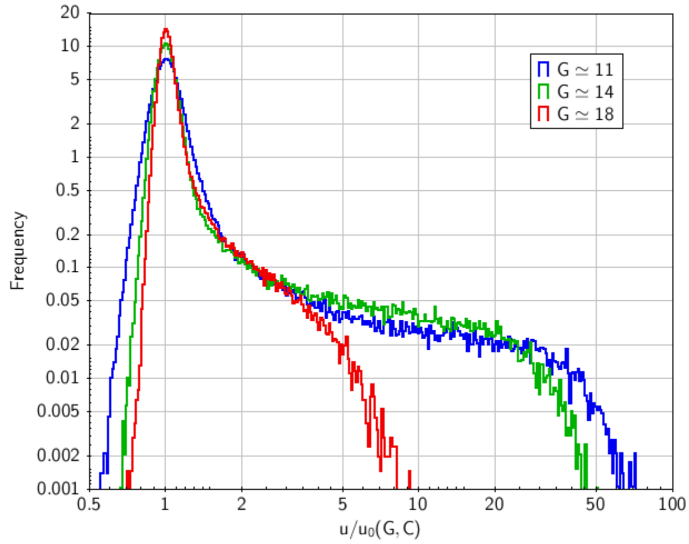

FIGURE 10: Same as Fig. 1 but for the RUWE ($u_{\rm norm} = u/u_0$).

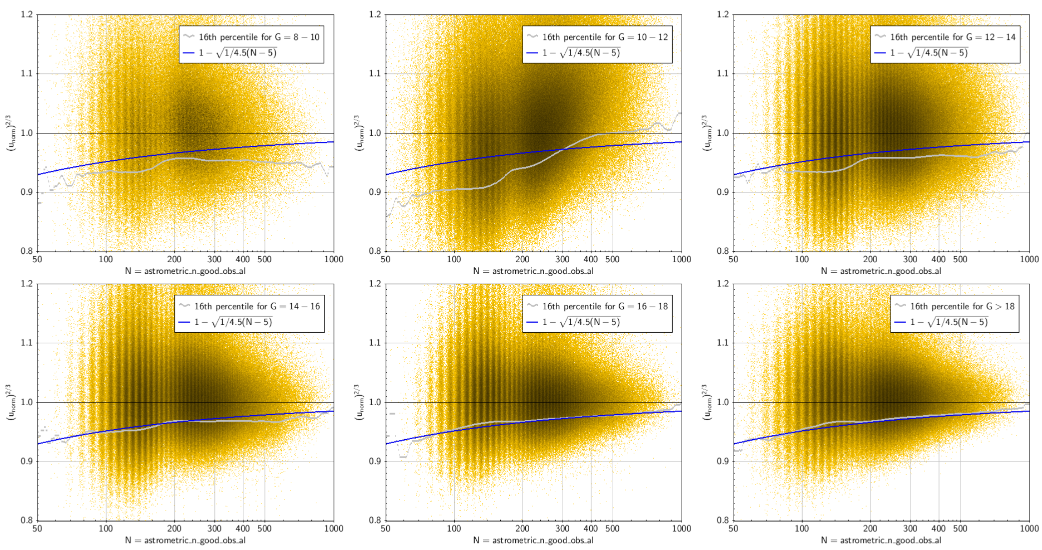

FIGURE 11: Distribution of the RUWE (to power $2/3$) as a function of the number of observations, N. The grey curve is the 16th percentile. The blue curve is the expected location of the 16th percentile for the theoretical distribution (see text).

5 The distribution of the RUWE

Figure 10 shows the distribution of the RUWE ($u_{\rm norm}$) for the same magnitude samples as in Fig. 1. As expected the peak is now located close to $u_{\rm norm} = 1.0$ for all magnitudes, and it is also slightly narrower, especially for $G \simeq 11$.

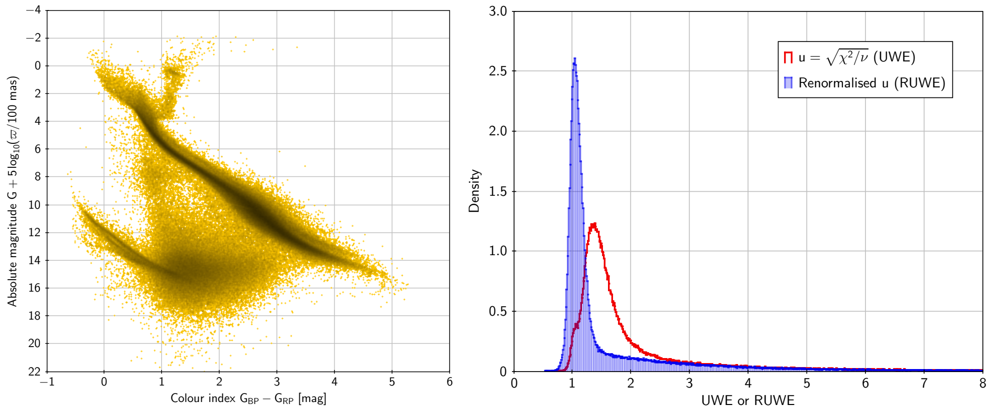

FIGURE 12: *Left:* HRD for a sample of 338 833 sources nominally within 100 pc of the Sun. *Right:* distributions of UWE and RUW for the same sources.

For well-behaved solutions one might expect that the re-normalised chi-square statistic $(N-5)\times\text{RUWE}^2$ approximately follows the chi-square distribution with $N-5$ degrees of freedom. To test whether this hold in practice is difficult because of the inevitable presence of binaries and other complications that create a long tail of large values of the RUWE. Even the main peak is expected to be distorted by the many solutions that are only moderately ill-behaved. However, the lower half of the distribution, essentially for RUWE $\lesssim 1.0$, should be relatively unaffected by these complications and could therefore sensibly be tested. This is done in Fig. 11, where RUWE$^{2/3}$ is plotted against N for several magnitude intervals. From the cube-root transformation (cf. Eq. 1 and Sect. 4) one expects the 16th percentile of RUWE$^{2/3}$ to vary as $1-(4.5(N-1))^{-1/2}$. To a good approximation this is indeed the case for the fainter sources ($G \gtrsim 16$), but not for the brighter ones. Especially in the interval $G = 10$–12 the distribution is quite different. A conclusion from this is that thresholds in RUWE should be set based on empirical evidence rather than theoretical distribution. An example is given in Sect. 6.

6 An example using the RUWE

The following application illustrates some of the benefits of using RUWE instead of UWE for selecting sources with ‘good’ astrometry. Figure 12 shows the Hertzsprung–Russell diagram (HRD) for 338 833 sources in *Gaia* DR2, nominally within 100 pc of the Sun, together with the distributions of UWE and RUWE for the sample. The sample is obtained with the criteria

$$\left.\begin{array}{rl} \text{(i)} & \varpi > 10 \text{ mas} \\ \text{(ii)} & \varpi/\sigma_\varpi > 10 \\ \text{(iii)} & \texttt{phot\_bp\_mean\_flux\_over\_error} > 10 \\ \text{(iv)} & \texttt{phot\_rp\_mean\_flux\_over\_error} > 10 \end{array}\right\} \quad (5)$$

This is the same as 'Selection A' in Appendix C of Lindegren et al. (2018), and as discussed in that reference the HRD is not at all as clean as could be expected from the rather stringent astrometric and photometric criteria. In particular many points fall between the main and white-dwarf sequences, where only few are expected, and in a big cloud roughly around colour index 1.5 and absolute magnitude 15.

The distributions of UWE and RUWE differ in the ways already discussed (Sect. 5), although much clearer now than when random samples are considered: the peak of RUWE is located approximately at 1.0 and is much narrower than for UWE. However, a more interesting difference concerns the shape: for RUWE there seems to be a clear breakpoint around RUWE $= 1.4$ between the expected distribution for well-behaved solutions and the long tail towards higher values. Although the long tail is also present in UWE, there is no clear breakpoint. Thus, looking at the distribution of RUWE it is quite natural to adopt RUWE ≤ 1.4 as a criterion for 'good' solutions. This retains 236 684 or 70% of the sources. The HRD for this subsample is shown in the upper-left panel of Fig. 13.

The distribution of the UWE (red histogram in Fig. 12) does not naturally suggest a value for the cut. but a fair comparison with the RUWE is obtained if the cut is selected to retain the same proportion of sources, i.e. 70%. This is obtained with the cut UWE < 1.96, resulting the the HRD in the upper right panel of Fig. 13. The HRD based on the cut in RUWE is definitely cleaner than the one based on a cut in UWE, for the same number of sources, but more important are the sources that are retained by the cut in RUWE but removed by the cut in UWE: these are shown in the lower left HRD. This includes most of the giants and a number of sources at the red end of the main sequence and the blue end of the white dwarf sequence. With reference to Fig. 6 this can easily be understood: these sources are either apparently very bright (the giants) or have rather extreme red or blue colours, and in all these cases the 'normal' UWE could easily exceed 1.96. The lower right diagram shows the sources retained by the cut in UWE but removed by the cut in RUWE. While this includes many sources with apparently good parallaxes and photometry, most of them ($\sim$85%) belong to the cloud far below the main sequence.

In summary, the selection based on the RUWE results in a sample that is both cleaner and more complete than comparable cuts based on the UWE. The use of the RUWE is especially useful for samples including very bright, very blue, or very red sources.

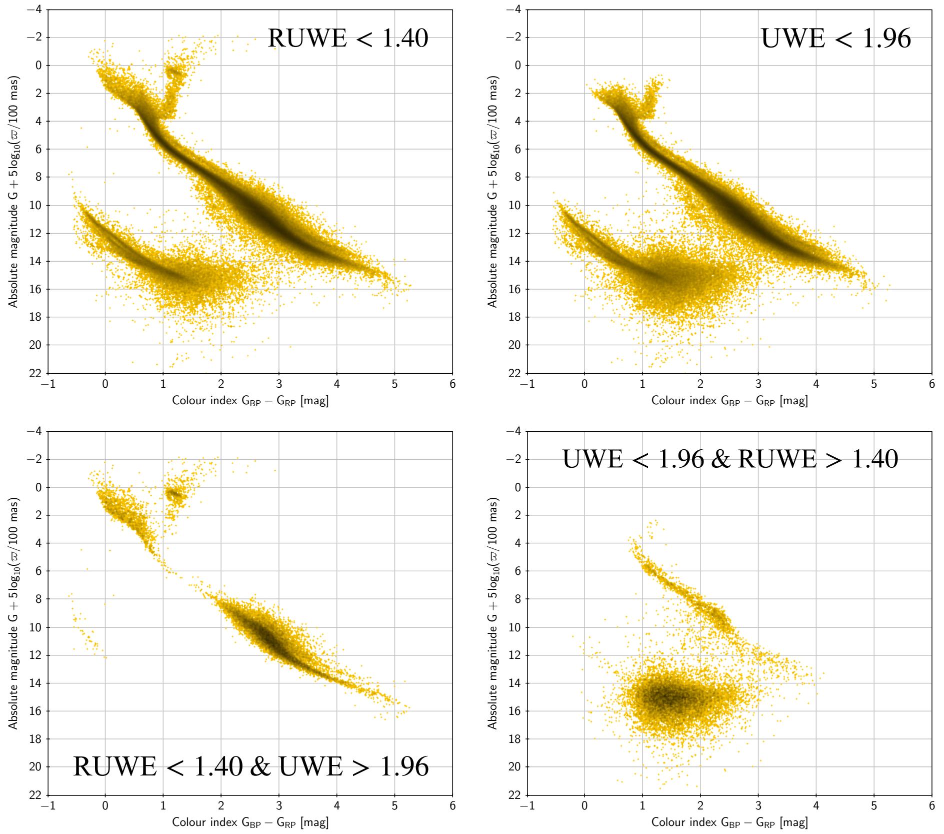

FIGURE 13: HRD for various subsamples of the sources in Fig. 12, selected according to UWE and RUWE as indicated in the diagrams. The upper diagrams show the selections obtained with a cut in RUWE or UWE, in both cases retaining 70& of the sources. The lower diagrams show the sources 'gained' by this cut compared with the other cut.

7 Tables of the reference UWE

The CSV file `table_u0_g_col.txt` contains a table of $u_0(G, C)$ for $G = 3.6(0.01)$ 21.0 mag and $C = -1(0.1)10$ mag. There is one line per value with the three columns G, C, and u_0. It is recommended that it is not used beyond these limits in G and C, and that linear interpolation in both coordinates are used to avoid discontinuities. (However, the sampling is dense enough that no serious error is made if the nearest tabulated values is used instead of interpolation.) The table is only valid for five-parameter solutions (`astrometric_params_solve` $= 31$) in *Gaia* DR2 with a valid colour index $C =$ `bp_rp`.

Gaia DR2 contains many valid five-parameter solutions without a colour index. For these, the function $u_0(G)$ tabulated in the CSV file `table_u0_g.txt` can be used instead. $u_0(G)$ was computed as the 41st percentile of u in each magnitude bin, using sources of all colours (or a random subset of them). In the same way as for $u_0(G, C)$, only sources with ‘good’ colours were used for the calculation of $u_0(G)$.

A third CSV file `table_u0_2D.txt` is also provided. This contains the same information as `table_u0_g.txt` and `table_u0_g_col.txt`, but merged in a single file and in a format that may be more convenient in some applications. See `readme.txt` for details.

8 Conclusions

For *Gaia* DR2 the UWE defined by Eq. (2) provides a more consistent measure of the astrometric goodness-of-fit than alternative statistics such as the excess noise. To be really useful, however, the UWE requires re-normalisation depending on the magnitude and colour of the source. Tables are provided to facilitate this. The usefulness of the resulting RUWE is evident in simple applications like the HRD of nearby stars.

References

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1

Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2

Wilson, E. B. & Hilferty, M. M. 1931, Proc. Natl. Acad. Sci. USA, 17, 684

Acronyms

The following table has been generated from the on-line Gaia acronym list:

Acronym	Description
CCD	Charge-Coupled Device
CSV	Comma-Separated Value (database output format, e.g., for MS Excel)
HRD	Hertzsprung-Russell Ddiagram
IAU	International Astronomical Union
ICRS	International Celestial Reference System
RAS	Rien à Signaler (nothing special to be noted)
RMS	Root-Mean-Square
RUWE	Re-normalised Unit-Weight Error
TN	Technical Note
UWE	Unit-Weight Error