Investigation of standard uncertainties in flux in EDR3.

prepared by: F. van Leeuwen
affiliation : Institute of Astronomy, Cambridge
approved by: D.W.Evans
reference: GAIA-C5-TN-IOA-FVL-184-1
issue: 1
revision: 0
date: Nov.10, 2021
status: Issued

Abstract

The standard uncertainties in flux measurements for broad-band photometry (G, BP, RP) as published in EDR3 are investigated. Expected uncertainty levels are derived from gate and saturation statistics and mean background contributions. A comparison with similar results for DR2 data is included. Jira issue C5PPV-79 is related to aspects described in this TN.

Document History

Issue	Revision	Date	Author	Comment
1	0	2021-11-10	FVL	Approved
D	8	2021-11-06	FVL	Suggestions from dwe applied, bibliography ipdated
D	7	2021-05-12	FVL	Added paragraph explaining when the hook appears
D	6	2021-04-23	FVL	Added one mre reference
D	5	2021-04-08	FVL	Completed final read through draft
D	4	2021-04-07	FVL	Added references and links, some adjustments of text
D	3	2020-07-29	FVL	A few more updates and corrections applied
D	2	2020-06-30	FVL	Numerous updates following comment DWE
D	1	2020-05-24	FVL	First draft completed
D	0	2020-05-01	FVL	Creation

Contents

1 Introduction

The Gaia photometry is of such high precision and consistency that some aspects of it can only be revealed through comparisons between different data releases. A similar approach is used to aid the investigations of the astrometric data (Lindegren et al., 2021). Such comparisons are presented here for a sample of around 3 million stars with parallax values greater than 4 mas in both EDR3 and DR2. By choosing this limit it is possible to cleanly separate different populations of stars and how these reflect in the comparisons. Amongst others, differences in brightness as a function of magnitude and colour are investigated, as well as standard uncertainty on mean fluxes.

The DR2 and EDR3 reductions of the Gaia photometry are largely independent (in particular for the G band as based on significant improvements in the Image Parameter Determination (IPD) (Rowell et al., 2021) and in both the calibration model and selection of the stars as used in PhotPipe (Riello et al., 2021). This can result in differences in non-linearity and reference passband between the two systems. The first would be observed as a correlation between brightness and difference in brightness between the systems. The second would be observed as a correlation between colour and difference in brightness, with additional possible dependencies on luminosity type.

The EDR3 photometry includes a number of new concepts in the data analysis. In IPD colour dependencies were taken into account for the first time. In PhotPipe the flux loss in AC windows, related to the AC position of the image in the window, was newly taken into account. The reference set of stars defining the system in EDR3 is also different and smaller in number than that used in DR2. Also the time span covered by the reference set is longer and relatively less affected by mirror contamination.

Supporting data from an earlier study on gating strategies and saturation in data segment 3 are also used (van Leeuwen, FVL-181), mainly to assess expectation values for standard uncertainties on fluxes and magnitude as affected by gates and saturation.

Some results from this study have also been reported as Jira C5PPV-79.

1.1 Objectives

The main objective of this study is to reach a better understanding of aspects of the Gaia broadband EDR3 photometry. This is the kind of investigation that may be triggered by finding significant systematic differences between DR2 data that was used before and the new EDR3 data. I consider it desirable that we show a good understanding of aspects exposed such comparison.

1.2 Context

This note is to contribute to the detailed understanding of properties of the photometric data published in EDR3.

2 The data

Data have been extracted from DR2 and EDR3, covering the entire sky for stars with parallax values greater than 4 mas and errors on the parallax less than 1 mas. This amounts to 5.4 million sources in EDR3 and 5.6 million sources in DR2. The two data sets were matched using the sourceId values. The number of false identifications obtained this way is extremely small, and not affecting any of the further discussions. The merged data set contained 3.7 million sources. The main loss came from small values of the parallax, causing different sources to be selected for the different archives. A small number of very bright sources could also not be matched, as these are not present in DR2.

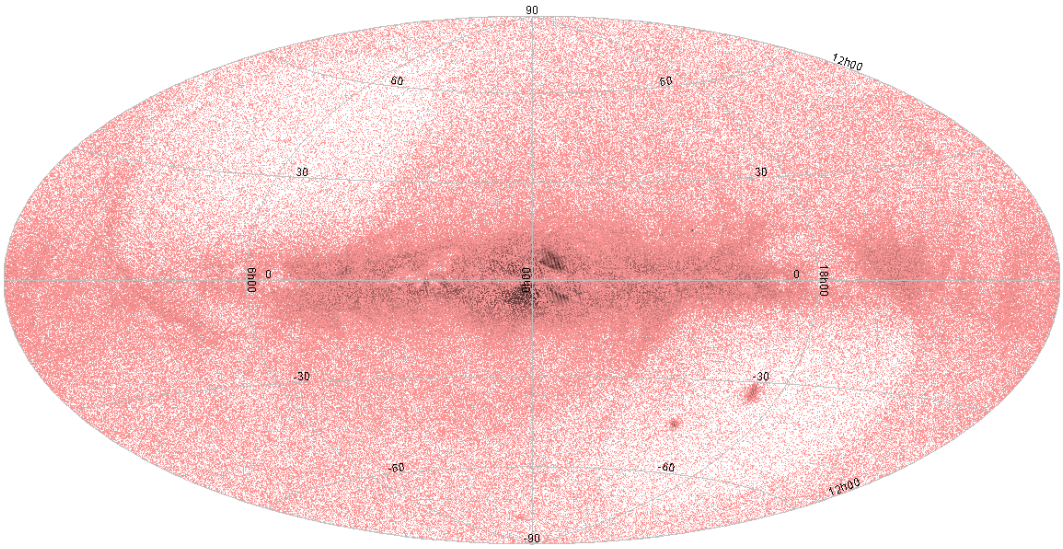

Figure 1: The distribution of the rejected sources on the sky, galactic coordinates.

The merged data set was further cleaned from any sources with `phot_proc_mode` not equal to zero in both releases, thus keeping only data indicated as "gold" in both data sets. This left 3.2 million sources. The rejected sources show a significant concentration towards the galactic centre, the LMC and SMC, as well as a more or less even distribution over rest of the sky (Figure 1).

The next main filter is the removal of stars with poor parallax determinations. These can be distinguished when comparing parallax error with G magnitude values, as shown in Fig. 2, top

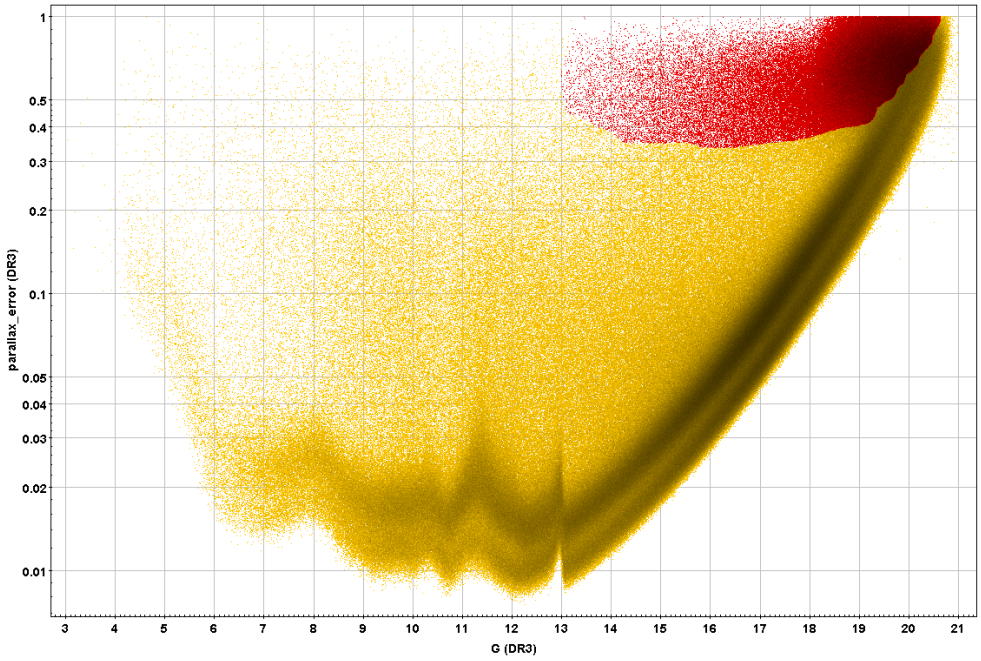

Figure 2: Parallax errors as a function of magnitude in the EDR3 data. Sources in the red are have been removed from further discussion. The two bands in the accepted data are related to the number of observations as determined largely by the position of a star in ecliptic latitude.

graph. The exact boundaries of this group can not be established, but the main purpose here is to remove the bulk of data that is very obviously wrong. A concentration of 291 thousand faint sources with relatively large errors on the parallax were removed, leaving 2.9 million sources for further examination. The largest concentrations of these sources is near the galactic plane, in particular towards the galactic centre, and the SMC and LMC (Fig. 1). The sources were removed as they could not provide a reliable classification in the HR diagram, which is required in the further analysis. For the remaining 2.9 million sources, around 96 per cent have relative parallax errors below 10 per cent, less than 0.1 per cent have relative errors above 20 per cent. This means that for nearly all the used sources the errors on the absolute magnitudes are close to symmetric and a direct reflection of the relative errors on the parallaxes.The relation between the uncertainty on the parallax ϖ (in mas) and the distance modulus D follows from:

$$\begin{aligned} D &= 10 - 5\log(\varpi) \\ D + \sigma D &= 10 - 5\log(\varpi(1 + \sigma\varpi/\varpi)) \\ &= 10 - 5(\ln(\varpi) + ln(1 + \sigma\varpi/\varpi))/\ln(10) \\ \sigma D &= -5(ln(1 + \sigma\varpi/\varpi))/\ln(10) \\ &\approx (-5/\ln(10))(\sigma\varpi/\varpi) \\ &= -2.17(\sigma\varpi/\varpi). \end{aligned} \quad (1)$$

The asymmetry on σD in this approximation is to first order

$$\delta\sigma D \approx 2.17(\sigma\varpi/\varpi)^2/2, \tag{2}$$

At $\sigma\varpi/\varpi < 0.1$ the asymmetry in the uncertainty σD is less than 5 per cent.

3 The HR diagram

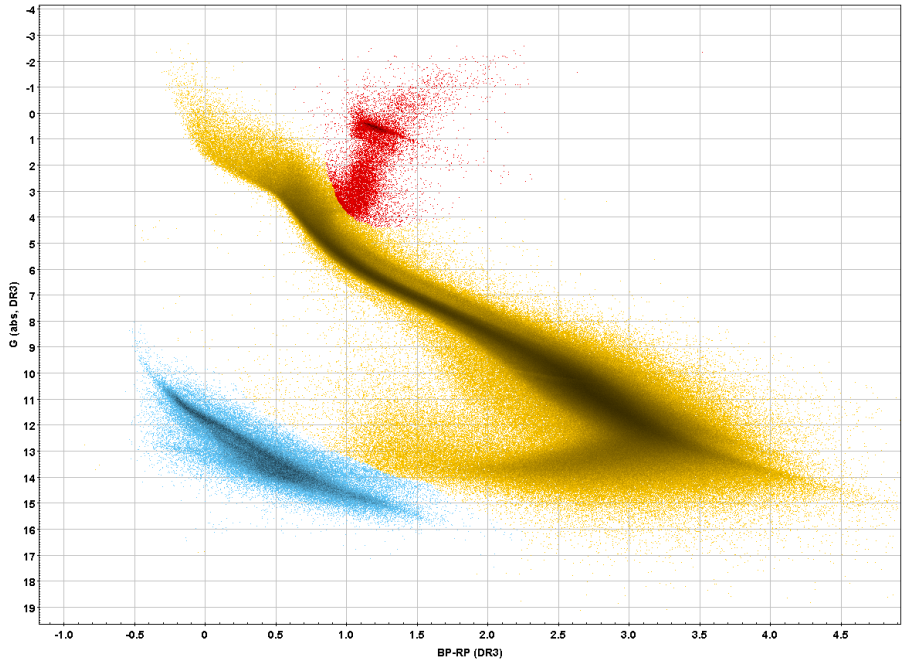

Figure 3: The HRD diagram for 2.9 million stars with parallaxes larger than 4 mas, and parallax errors less than 1 mas. Stars classified as main sequence objects are coloured ochre, White Dwarfs light blue and Giants red.

The HR diagram (Fig. 3) was produced for the EDR3 data only, to distinguish between White Dwarfs, Main Sequence stars and Red Giants, which, because of their different spectral characteristics, may react differently on small passband changes. The separation of the 99 852 White Dwarfs (a factor 3 increase from DR2) and 15 876 red giants follows easily from the characteristics of the HR diagram. The remaining 2 803 096 stars are considered to be main sequence stars. The main sequence shows a ”hook” similar to what was observed in DR2. This ”hook” will be analysed further in Section 9. The selection made here ensures that we have good populations of blue and red stars with very different brightness at the same colour index. This is helpful when examining passband and linearity changes. A lower value of the parallax cut-off would not help the White Dwarf distribution and would introduce more reddened bright stars, but it could have helped extending the population of RGB stars.

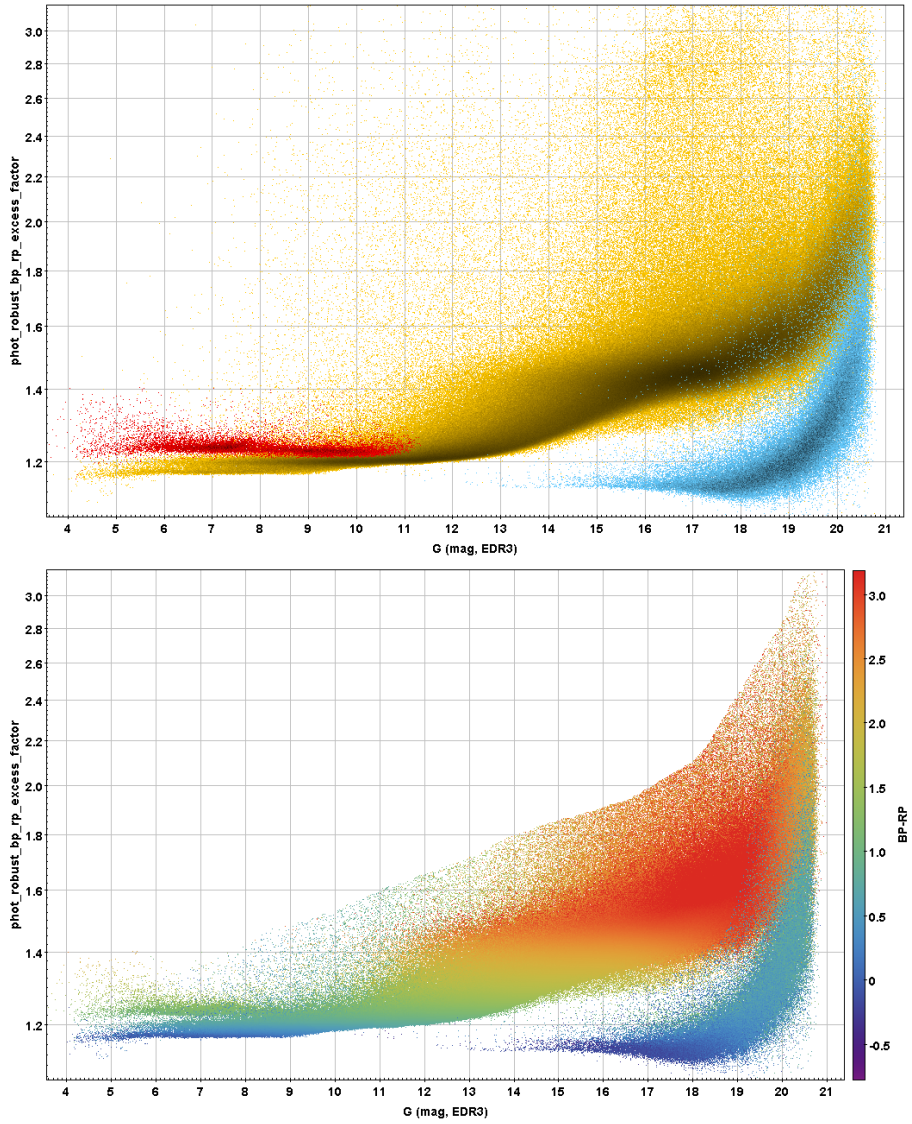

Figure 4: The BP, RP excess factor in the EDR3 data. Top: as seen for the three different luminosity selections (colours as in Fig. 3). Bottom: with dependencies on BP−RP colour index after removing outliers.

The data was further cleaned by examining the `phot_bp_rp_excess_factor` in the EDR3 data. This was done per luminosity group, as it can be seen clearly in Fig. 4 (top) that different selection criteria should be applied to each group. This left 98 121 White Dwarfs, 15 702 red giants and 2 768 688 main sequence stars.

4 Mean magnitudes and errors

The fluxes as measured are considered to be random samples of Poisson distribution for the "volume" of the image, as recorded in ADU. The mean flux expectation value $F_0(\mathrm{ADU})$ is approximated by the observed ADU flux $F_i(\mathrm{ADU})$:

$$F_i(\mathrm{ADU}) = F_0(\mathrm{ADU}) + \epsilon_i(\mathrm{ADU}),\ \left\langle \epsilon_i^2(\mathrm{ADU}) \right\rangle = F_0(\mathrm{ADU}). \tag{3}$$

The gain factor from ADU to $e^- s^{-1}$ is given as $f = 0.2566$, and transforms the observed flux in ADU to flux in $e^- s^{-1}$:

$$I_i = F_i(\mathrm{ADU})/f. \tag{4}$$

Expressing Equ. 3 in flux in $e^- s^{-1}$ gives:

$$I_i = I_0 + \epsilon_i,\ \left\langle \epsilon_i^2 \right\rangle = f \cdot I_0 = 0.2566 \cdot I_0, \tag{5}$$

Figure 5 shows how the IPD errors on fluxes very closely follow the Poisson noise expectation.

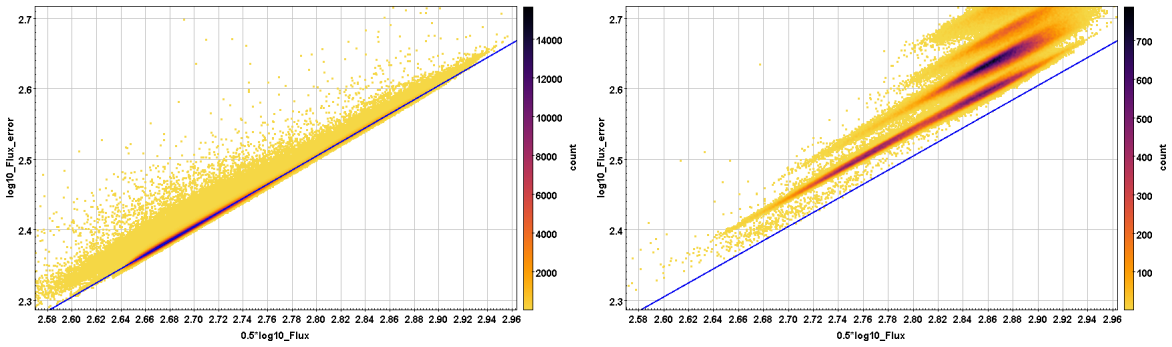

Figure 5: Relation between flux and error on flux for no-gate observations without saturation (left) and for images with one or more saturated pixels (right). The diagonal blue line represents the expected photon noise for not-saturated images. Data for AF2, Row 4, both fields of view.

For images with saturated pixels this remains true, but for these the "image volume" available for fitting has been reduced. It should be noted here that at least for gated observations the gain factor appears to smaller (close to 0.2400), which still remains unexplained.

For faint stars, the dispersion of observed flux with respect to the mean flux value is set by the expectation value for the mean flux plus the flux from the background. This applies to individual observations as well as the accumulated flux. The background contribution amounts

to the background per pixel (or sample) times the surface area in pixels (or samples) used for the image parameter determination.

Stars have been observed by differing numbers of times, and with various gate settings. The different gate settings used for G band are shown in Table 1.

Gate	Pixels	g
0	4494	1.0
12	2900	0.645
11	2048	0.456
10	1024	0.228
9	512	0.114
8	256	0.057
7	128	0.028
4	16	0.0035

Table 1: Gates as deployed in the G band, the integration length in pixels, and normalized to a full-length integration.

Two different solutions are considered. First for faint stars, for which the background contribution plays an important role:

$$I_i = I_0 + \epsilon_i, \ \left\langle \epsilon_i^2 \right\rangle \approx f(I_0 + < b_i >), \tag{6}$$

For the accumulated data the total flux gathered is:

$$I_T = \Sigma I_i, \ \left\langle \epsilon^2 \right\rangle = f\Sigma(I_0 + b_i). \tag{7}$$

The signal over noise, given n observations, is then given by:

$$I_T/\sigma I_T = n \cdot I_0/\sqrt{n(f(I_0 + \langle b \rangle))} = \sqrt{n} I_0, /\sqrt{f(I_0 + \langle b \rangle)} \tag{8}$$

where $I_0 = \langle I_i \rangle = \Sigma I_i/n$ and $\langle b \rangle = \Sigma b_i/n$. The mean error per observation is given by:

$$\sigma_I = \sqrt{f(I_o + \langle b \rangle)}, \tag{9}$$

and the mean signal over noise per CCD observation:

$$I/\sigma_I = I_0/\sqrt{f(I_o + \langle b \rangle)}. \tag{10}$$

The second solution applies to bright stars for which a mixture of gate settings may have been used. The observed flux is then given by:

$$I_i = g_i I_0 + \epsilon_i, \ \left\langle \epsilon^2 \right\rangle = f \cdot g_i I_0. \tag{11}$$

Accumulated are, however, values of I_i/g_i:

$$I_i/g_i = I_0 + \epsilon'_i, \ \left\langle \epsilon'^2_i \right\rangle = f \cdot I_0/g_i. \tag{12}$$

In the unit-weight solution for gated observations the error on the mean flux is given by the inverse square root of the sum of the weights $g_i/(f \cdot I_0)$:

$$\sigma_{I_0} = \sqrt{f \cdot I_0}/\sqrt{\Sigma g_i} = \sqrt{f \cdot I_0}/\sqrt{n \langle g_i \rangle}, \tag{13}$$

where the mean gate $\langle g \rangle = \Sigma g_i/n$ There is no significant overlap between these two situations. When gates are applied, the background tends to be a negligible contribution to the total flux. The signal over noise for gated observations is:

$$I_T/\sigma_{I_T} = \sqrt{n \langle g \rangle}\sqrt{I_0/f}, \tag{14}$$

and the mean signal over noise per observation is:

$$I/\sigma_I = \sqrt{\langle g \rangle}\sqrt{I_0/f}, \tag{15}$$

Statistics on the mean gate as a function of the on-board G magnitude estimate have been

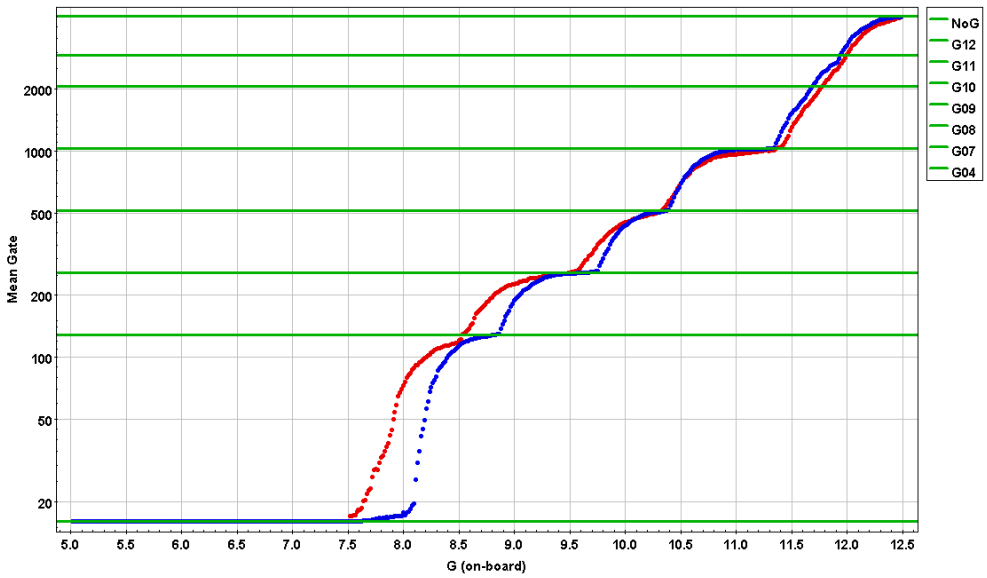

Figure 6: The mean gate values as a function of on-board G magnitude. Blue: data before OBMT rev.4172, Red: data after OBMT rev.4205. Also shown are the gate positions.

accumulated for all of AstroElementaries of data segment 3 (van Leeuwen, FVL-181), for stars with G brighter than 12.7. Around OBMT rev.4200 the gating criteria were updated, and data for the two regimes have been accumulated separately (see Fig. 6). The actual values used in further calculations have been normalized to the no-gate level, using the g values in Table 1.

The expressions in Equ. 8 and 14 are, to a good approximation, the equivalent of the `phot_g_mean_flux_over_error` values provided by the archive. By dividing these values by the square root of `phot_g_n_obs` , normalized signal-over-noise values are obtained independent of the variations in the number of observations over different sources.

The way these relations are affected by saturation depends very much on how saturations is handled in IPD. Generally it means that a reduced number of pixels is available for IPD, reducing the photon count of the usable image and thus decreasing the signal-to-noise level for the recovered image flux in comparison with not-saturated images of the same source.

5 Magnitude comparisons, the G band

A magnitude dependence of differences between magnitudes as extracted from DR2 and EDR3 indicates a difference in the linearity between the two systems. It does not define the magnitude of the non-linearity in either system, for that a reliable third-party system is required. This may be difficult to obtain at the precision level of the Gaia photometry. Comparisons with respect to SPSS and CALSPEC give a good indication of the relative magnitude terms present between the different data sets.

If we have a linear distortion on the intensity scale, $I = I_0 \cdot (1 + \epsilon)$, then this affects the reproduced magnitudes by:

$$\begin{aligned} m_s &= m0 - 2.5\log_{10}(I_0^{1+\epsilon}) \\ &= m0 - 2.5\log_{10}(I_0) - 2.5\epsilon\log_{10}(I_0) \\ &= m_r + \epsilon(m_r - m_0), \end{aligned} \tag{16}$$

where m_s is the magnitude in the photometric system and m_r the correct magnitude in a linear system.

Figure 7 shows the observations (left) for the three luminosity groups. The slope has been determined based on the main sequence stars between magnitudes 6 and 18. For brighter stars there are significant further distortions due to differences in the handling of saturation, for fainter stars the noise level and further systematics, related to the ”hook” in the HRD, cause problems. The fit to the data is obtained by:

$$G_{\rm DR2} = G_{\rm DR3} + (0.0043 \pm 0.0001)(G_{\rm DR3} - 12.0). \tag{17}$$

The equivalent correction, if applied just to DR2 data, is:

$$G'_{\rm DR2} = (0.9957 \pm 0.0001)G_{\rm DR2} + (constant), \tag{18}$$

which is close to the factor of 0.9966 (no error provided) derived by Casagrande & VandenBerg (2018) for the DR2 data compared to synthetic magnitudes. It may be possible to find out by

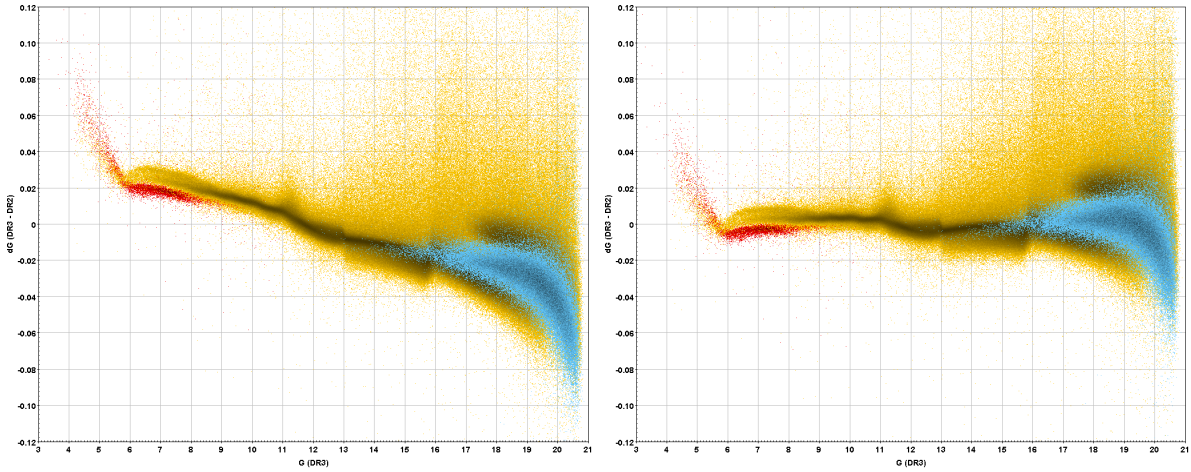

Figure 7: The G magnitude differences between EDR3 and DR2 as a function of the G magnitudes in EDR3. Left: As observed; right: after applying correction (see text). Colours as in Fig. 3.

means of the SPSS stars what the contributions from DR2 and EDR3 are to this difference, as at magnitude 6 the differences amount to 0.026 magnitudes. A similar distortion from the linear scale may also be reflected in the absolute passband calibration by Paolo Montegriffo. In his slides (here) at CU5M18 (slides 13 and 14) there is also an indication of a dependence, at a similar slope, on the G magnitude in the DR2 data, when compared with SPSS stars.

The data displayed in Fig. 7 show that at the faint end the White Dwarfs and faint red main sequence stars react in the same way, which is a strong indication that this is not primarily a passband adjustment issue. Also clearly visible in the same figure are the changes in the connections between different calibration units, in particular at G=16 for the transition between short and long 1D windows, at G=13 for the transition between 1D and 2D windows, and G=11.3 for the activation of gate 12 in 2D windows. The transition at G=16 amounts to about 0.005 magn for red stars, but seems to be much smaller for the White Dwarfs. The transition at G=11 is most likely affected by differences in saturation treatment between DR2 and EDR3.

The systematics of differences in G magnitude between DR2 and EDR3 as a function of the colour index should reflect any passband differences between the two systems. Figure 8 shows the data as observed, and after applying the correction given in Equ. 17. Although there are differences in the relation between colour and brightness depending on spectral characteristics, such differences as caused by a very small passband correction would be very much second-order effects. What is observed here is that the correction for differential non-linearity in the G band between the two system has, when applied to the colour dependencies, also largely removed discrepant behaviour for blue stars and the red giants. The remaining dependencies with colour are small, and show the typical third-order dependency on colour that is often observed in comparisons between photometric systems. The EDR3 and DR2 photometric systems are

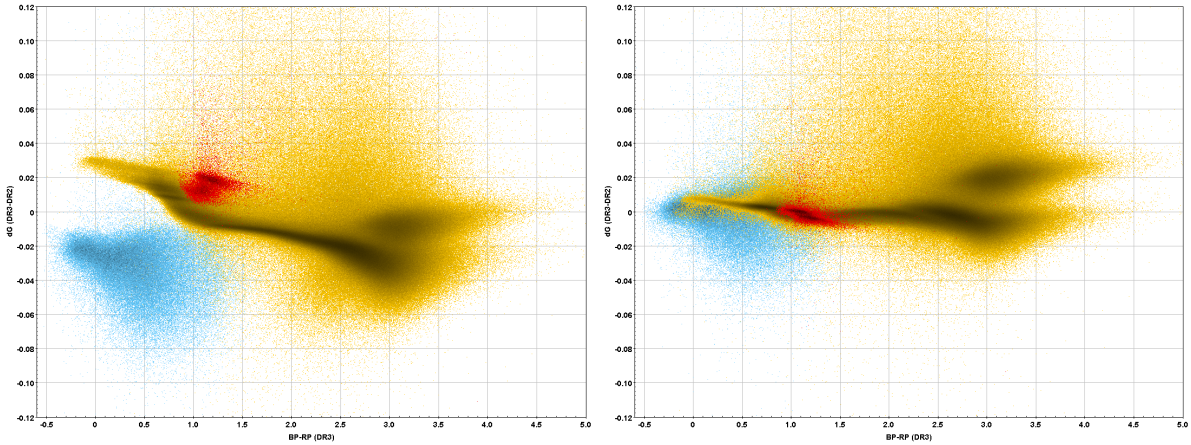

Figure 8: The dependence on colour index of the differences in G magnitude between EDR3 and DR2. Left: as observed; right: after applying the correction of Equ. 17 to the G magnitude differences. Colours as in Fig. 3.

slightly different. The additional concentration in the distribution beyond BP−RP$= 2.5$ will be referred to as the ”blob”. It contains around 0.5 million stars, and is described further in section 9.

6 G magnitude standard uncertainty levels

A first estimate of expected uncertainty levels in the Gaia broad-band photometry was provided in Jordi et al. (CJ-047). In it the effect of gates was assumed to be exactly limited to pre-mission specified magnitude ranges, which then provides a clean saw-tooth pattern of expected accuracies as a function of magnitude. In reality the gate activation is significantly smeared out due to the uncertainty in the onboard magnitude estimates and the adjustments of gate activation as a function of AC coordinate and CCD. The current study gives an assessment of the combined effects of photon noise, calibration noise and gate activation statistics on the uncertainty levels for individual transits as observed in cycle 3 data.

The standard uncertainty on the G fluxes is set by three contributions:

1. Photon noise, as described in Section 4, not including saturated pixels;
2. background contributions, adding to the photon noise but not to the flux of the source;
3. calibration noise, limitations in the calibration model to represent the data at sufficient accuracy.

In the data file are included the values of `phot_g_mean_fux_over_error` ($= I_0/\sigma_{I_0}$), and the number of observations contributing to the mean, `phot_g_n_obs` (n). Together they provide an estimate of the mean accuracy per observation (see Section 4):

$$\sigma_I/I = \sqrt{n}\sigma_{I_0}/I_0. \tag{19}$$

Figure 9 shows these data for EDR3. Following Equ. 10, the data for un-gated stars have been fitted with:

$$\sigma_I/I = \sqrt{0.45 * (I_0 + b)}/I_0, \tag{20}$$

as shown by the thick red line in Fig. 9, and with the background $b = 150$ e$^-$. This background level represents the total background flux contribution to the image, which for the faintest stars covers 6 AC compressed samples. The equivalent per sample is therefore $b = 25$ e$^-$. At pixel level, considering 12 pixels per sample, this would be $b = 2.0$ e$^-$ or about 0.5 ADU. The factor $f = 0.45$ is larger than the expected value of 0.2566, but that value could only be verified on the IPD data for 2D images. It should be noted that this is different from the effect of calibration noise, which would be contribution to the noise level that is independent of the flux value.

Also shown in Fig. 9 is information concerning window and gate activations, and saturation (as derived in van Leeuwen (FVL-181)), as well as the photon-statistics limited accuracy for the 1D window observations, which includes an average background contribution per image of 126 in flux.

The coverage is for data-segment 3, during which the on-board strategy for gate allocations was changed. This is reflected in a set of blue data points representing data from before the change, and green data points after the change. The blue data points represent roughly 3/4 of the mission, the green data points 1/4. Those data were determined with respect to the on-board magnitude estimates. To obtain them with respect to the calibrated G magnitudes will result in a further convolution with a Gaussian distribution with a width of a few times 0.1 magn, further smoothing out any features.

Some interesting aspects can be observed in Fig. 9, such as the increase in noise levels in the transition from 1D to 2D windows (thought to result from an issue in the 2D PSF models) and the apparent anti-correlation between average accuracy and the mean saturation level, in particular at G=11.4. For data observed with gate 4 the measurements are more homogeneous, and relative noise levels very similar to the 1D data are obtained until the effects of saturation become dominant.

Figure 10 shows the ratio (EDR3/DR2) between mean errors on the G magnitude for individual transits. As can be seen in Fig. 10, the results for the brightest stars have improved, but for stars observed in non-gated 2D windows there is mixed success of a kind that can probably be attributed to a change or adjustment of the calibration model. This caused the data near the window transitions to have improved, but possibly at some cost to the data halfway those transitions.

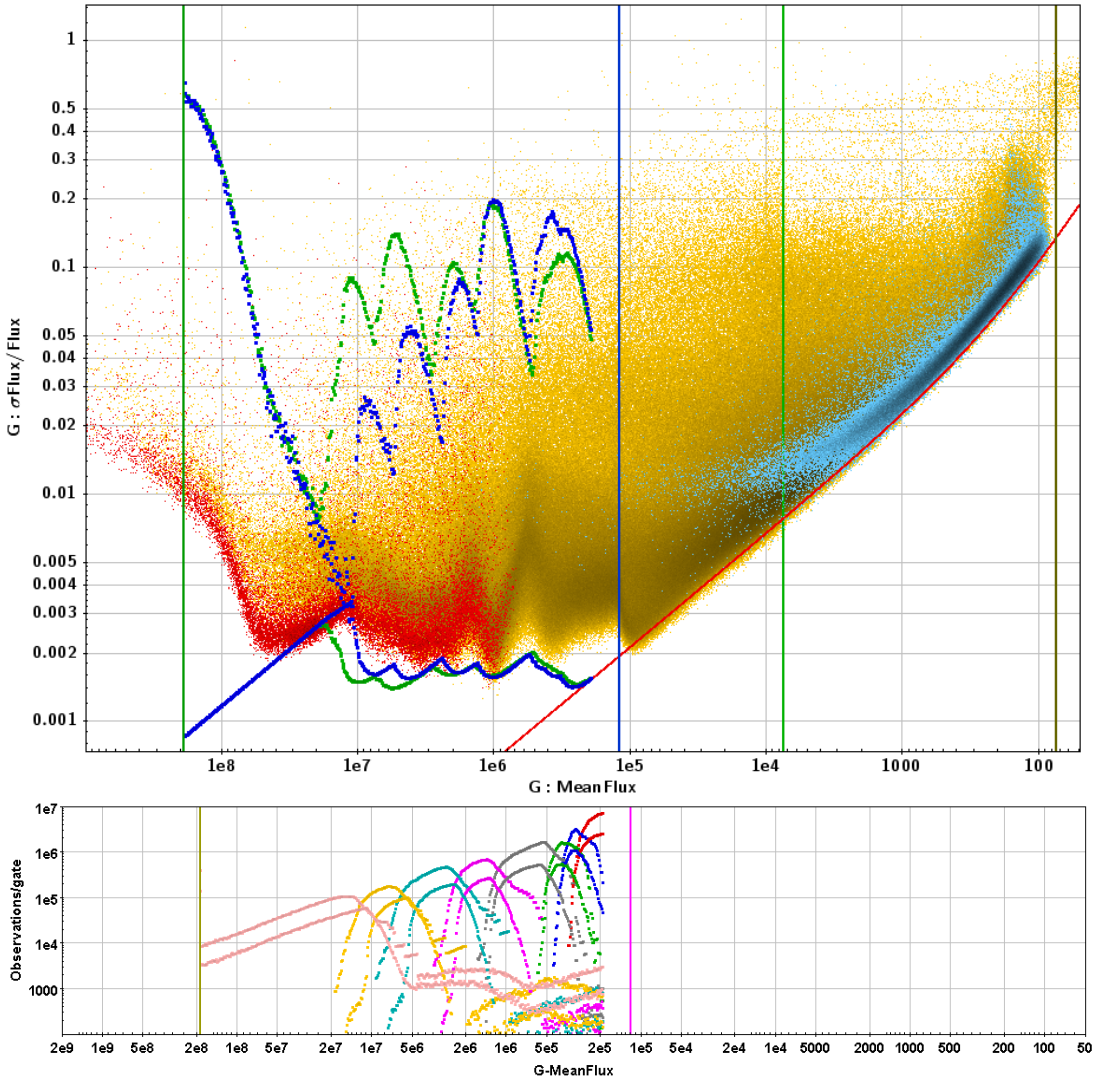

Figure 9: Photometric uncertainties on single observations in EDR3. Top: Standard relative uncertainty on G flux as a function of G mean flux values. Colours of the data points as in Fig. 3. Red diagonal line: Photon noise level for no-gate, corrected for an average background contribution; blue and green wiggly lines: Photon noise level based on mean gate. Vertical lines, from left to right: G=5, G=13 (transition from 1D to 2D windows), G=16 (transition from long to short 1D windows), G=21. Blue and green dots: 0.8 times the fraction of measurements with at least one saturated pixel. Bottom graph: numbers of observations for the different gate settings. from left to right: gates 4, 7, 8, 9, 10, 11, 12 and no-gate. Top curves: after OBMT rev.4205, bottom curves: before OBMT rev.4172.

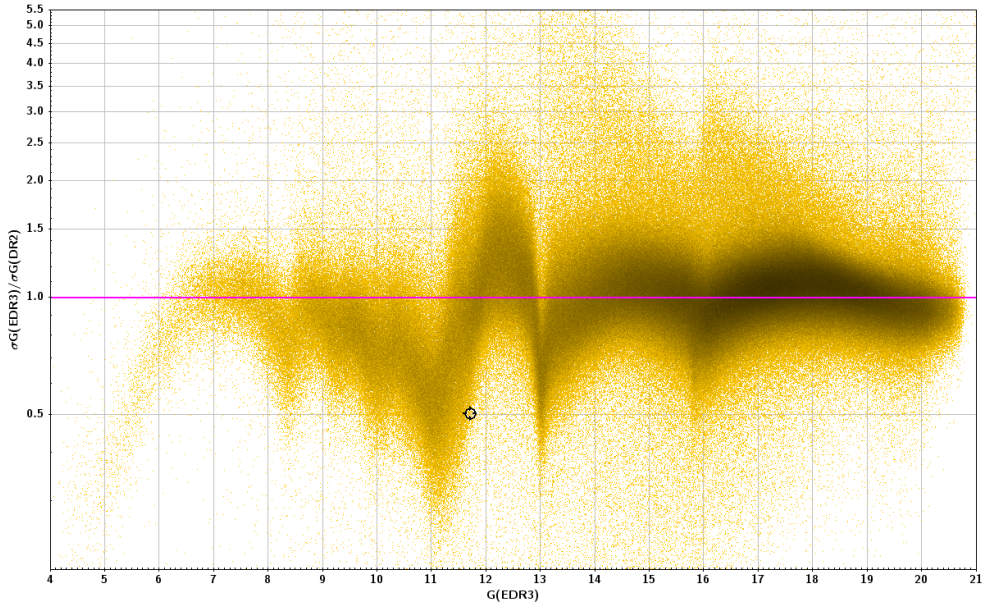

Figure 10: The ratio of the mean standard deviation on individual measurements between the EDR3 and DR2 photometry in the G band. Main sequence stars used only. Where the ratio is above 1.0 DR2 data performed better, where it is below 1.0 the advantage is with DR3 data.

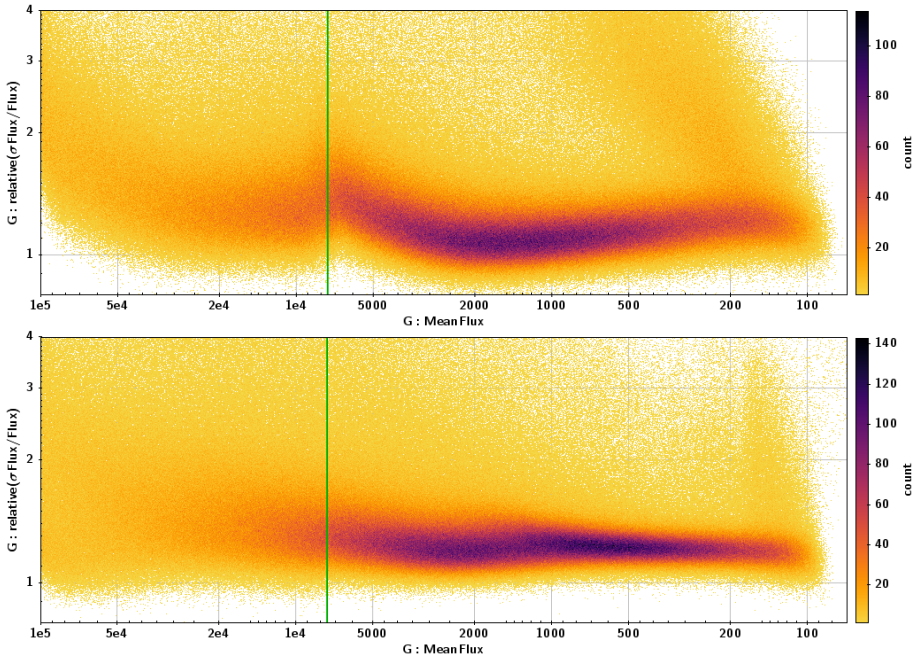

Figure 11: Ratio between the observed and modelled standard uncertainty dependencies for stars fainter than G magnitude 13.0. Top: for DR2, bottom: for EDR3. The vertical line represents the G=16 boundary between short and long 1D windows.

Figure 11 compares for DR2 and EDR3 data the standard uncertainties with the fit of Equ. 20. Here EDR3 data is clearly performing better than DR2 data, and features in the differences observed in Fig. 10 for these stars can be largely attributed to DR2 data.

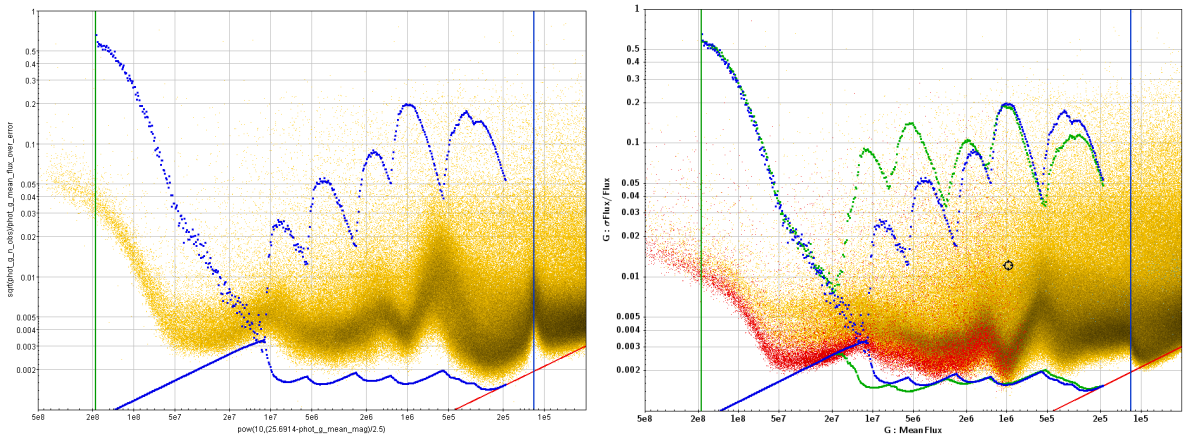

Figure 12: Mean uncertainties per CCD transit for gated data in DR2 (left) and EDR3 (right). The blue and green lines and points as in Fig. 9. The vertical lines refer to magnitude 5 and 13.

Figure 12 compares between DR2 and EDR3 the mean accuracies per CCD transit in the G band for mostly 2D images. There are clear overall improvements, such as for the saturated brightest stars and the transition from 1D to 2D windows. There remains an apparent anomaly in the relation between the relative number of stars affected by saturation and the mean noise level on the flux, in particular at G=11.4 (low saturation count, high noise level) and G=10.7 (high saturation count, low noise level). A further investigation of the epoch photometric data is needed to understand what the background is of this behaviour in the reduced data.

7 Magnitude comparisons, the BP and RP bands

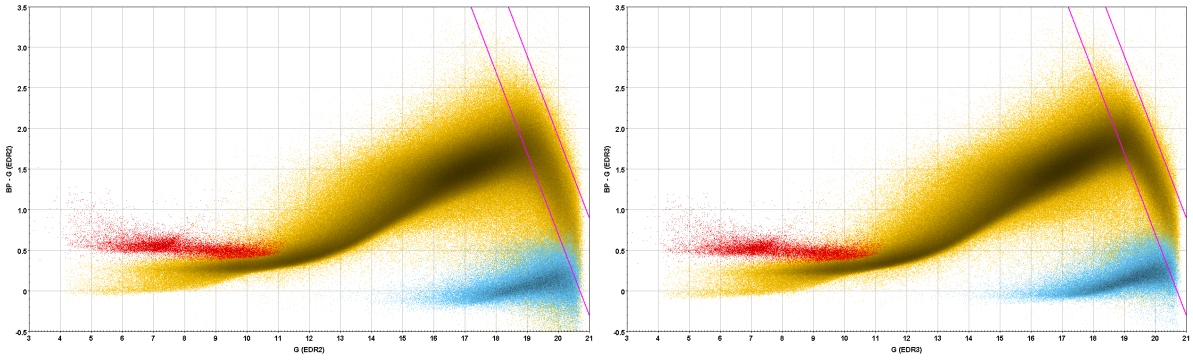

Figure 13: Comparison between G and BP magnitudes for DR2 (left) and EDR3 (right). The two diagonal lines are at BP=20.7 and BP=21.9. Colours as in Fig. 3.

Figure 13 shows the comparison between BP and G magnitudes in DR2 and EDR3. The data between the two diagonal lines in Fig. 13 are for stars for which the BP magnitude would be fainter than about G=21, at which point there is no longer enough signal to derive a reliable flux estimate. These stars are not marked directly in the data and can only be distinguished based on the ratio of the number of observations used for the BP and RP (see Section 9). This situation has not significantly changed between DR2 and EDR3, except for a gain of 0.2 to 0.3 magnitudes in the faint-limit for which sensible BP magnitudes are obtained. This feature is the same as the ”hook” in the HR diagram in Fig. 3 (see further Section 9). It creates a distortion of the colour indices for faint red stars, giving BP−RP values ranging from 1.5 to 2.5 when the actual values are probably more in the range 3.5 to 5. This may well affect (directly or indirectly) the photometric and astrometric calibrations at various stages: construction of the ELSF, IPD, astrometric calibrations, photometric calibrations, photometric passband reconstruction.

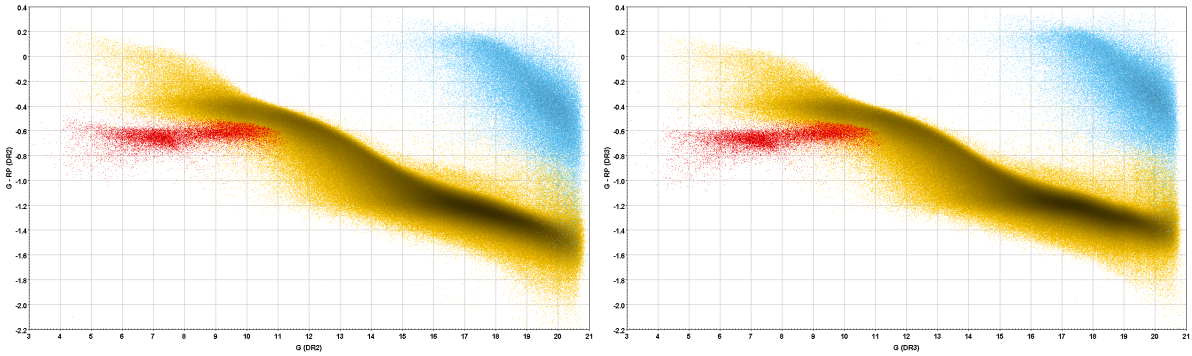

Figure 14: Comparison between G and RP magnitudes for DR2 (left) and EDR3 (right). Colours as in Fig. 3

Figure 14 shows the comparison between RP and G magnitudes in DR2 and EDR3. The differences are very small, and no magnitude cut-off is seen.

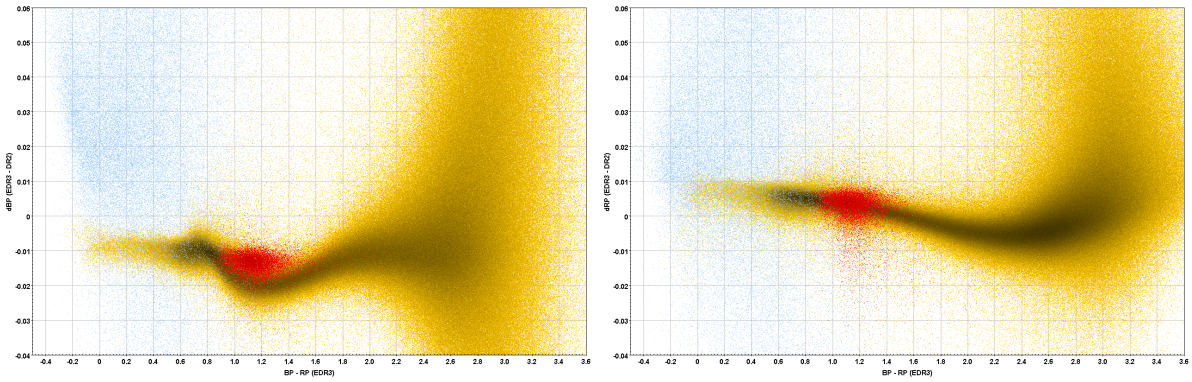

Figure 15: Small changes from DR2 to EDR3 in the effective passband are reflected in systematic changes in the BP and RP magnitudes as a function of colour index. Left: for BP, right: for RP . Colours as in Fig. 3

Small changes in the reference passband can be observed in both RP and BP when plotting the magnitude differences as a function of the colour index (see Fig. 15). It is noted that in the BP responses there is an average brightness increase of about 0.01 magnitudes, except for White Dwarfs, which seem to appear slightly fainter. In BP there is in addition a significant difference in the changes for giants compared to main sequence stars. In RP there is no significant systematic offset but only a smooth trend.

8 BP and RP standard uncertainties

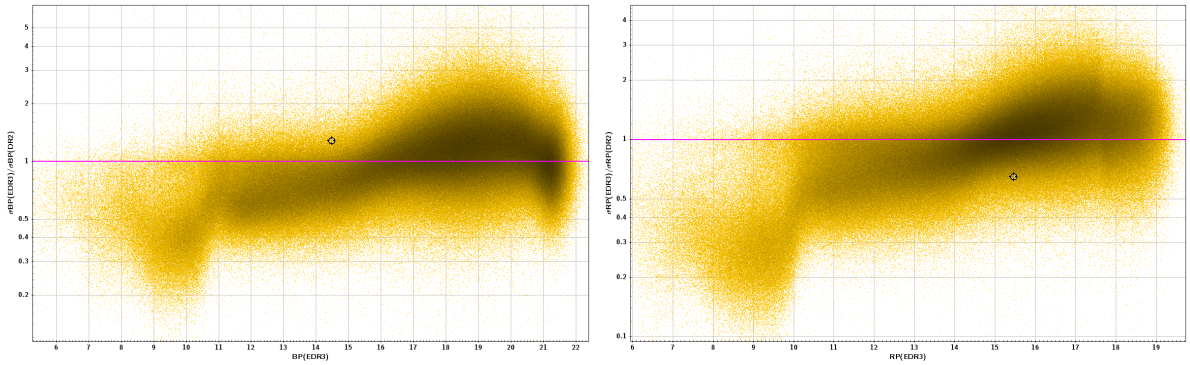

Figure 16: Error ratios EDR3/DR2 for BP (left) and RP (right) as a function of BP and RP magnitude respectively.

Figure 16 shows the error ratios for the "mean error per transit" in DR2 and EDR3 as a function of magnitude. The split between 1D and 2D windows is around BP≈11 and GP≈10, which also shows as a marked change in the error relations between DR2 and EDR3, with the standard uncertainties in the EDR3 analysis on average about four times smaller. It is not clear how this improvement was obtained.

Figure 17 shows the per transit average values of σ_I/I for BP and RP. For non-gated 1D data the thick red lines represent the photon statistics, including the contribution of the average background, following Equ. 10:

$$\begin{aligned} \sigma_I/I(\mathrm{BP}) &= \sqrt{0.28(I_{BP,0}+2300)}/I_{BP,0} \\ \sigma I/I(\mathrm{RP}) &= \sqrt{0.26(I_{RP,0}+1400)}/I_{RP,0} \end{aligned} \tag{21}$$

The mean background contribution per CCD transit is equivalent to 4.8e$^-$ and 2.9ee^-s^{-1} per pixel in BP and RP respectively. The scaling factor is in both case much closer to the expected level of 0.2566 than what is observed for the G band. No data is currently available on the average gating as a function of magnitude, which would allow a comparison between expected and observed noise levels for brighter stars.

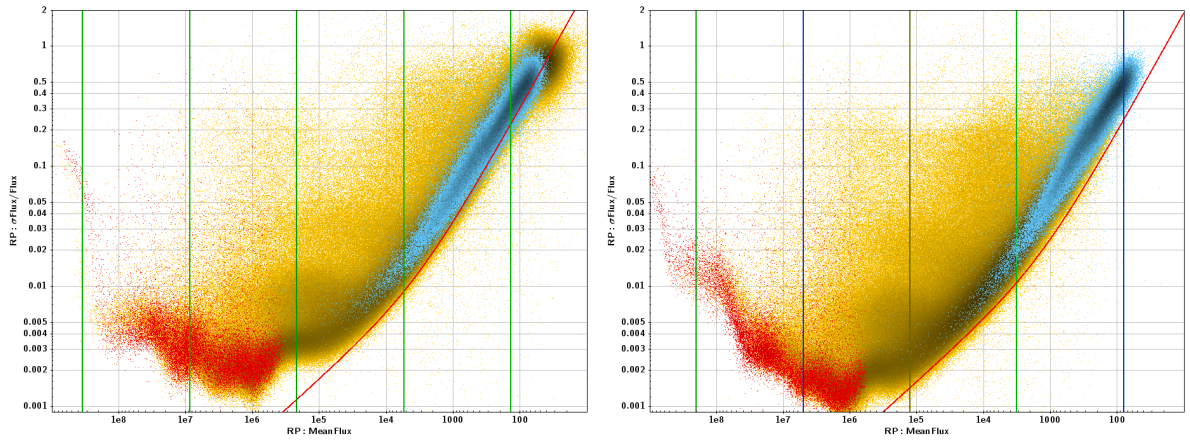

Figure 17: The relative uncertainty in flux per transit in BP (left) and RP (right) as a function of mean flux. The vertical lines represent from left to right magnitudes 4, 8, 12, 16 and 20. The thick red lines represent the photon statistics, including a background (see text). Ochre: main sequence stars, blue: White Dwarfs, red: Red Giants.

For a single observation a signal-to-noise level of 1.0 is reached at an average flux of 38.1, corresponding to a BP magnitude of 21.4. In the DR3 data most sources have between 20 and 60 BP observations, giving an expected signal-to-noise ratio at BP=21.4 of 4.7 to 7.7. An average signal-to-noise ratio of 3 can be expected down to BP=21.9 to 22.4. For cycle 4, going from 34 to 66 months of data, an average of 40 to 120 BP observations per source can be expected, extending the 3σ range to BP=22.2 to 22.8. The variation of accuracy will increase with lower fluxes because of the increasing dominance of the varying background contributions.

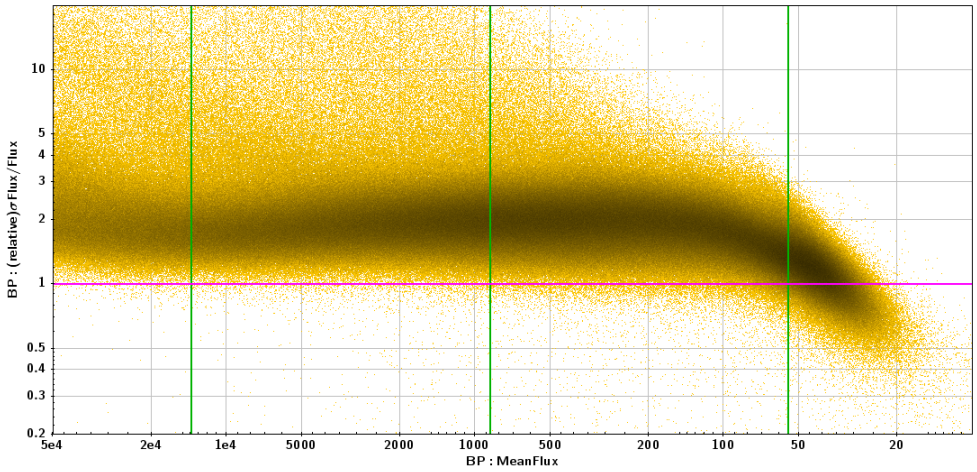

Figure 18: The ratio between the mean relative error on the flux and the photon noise model (thick line in Fig. 17). The green vertical lines represent, from left to right, BP magnitudes 15, 18 and 21. points: Main sequence stars only.

Figure 18 shows the ratio between the per CCD mean relative error on the BP flux and the

photon noise model as a function of mean flux in BP. At the faint end of the distribution the ratios become anomalous, showing errors systematically as low as 0.2 times the expected photon noise, a phenomenon associated with the ”hook” referred to above and in Section 9. Figure 18 appears to show that the bias in the B magnitudes becomes noticeable already at a mean flux of around 90, which seems high considering the flux cut-off value of 1 used in the reductions. It is still unclear why this should be the case.

9 The hook and blob

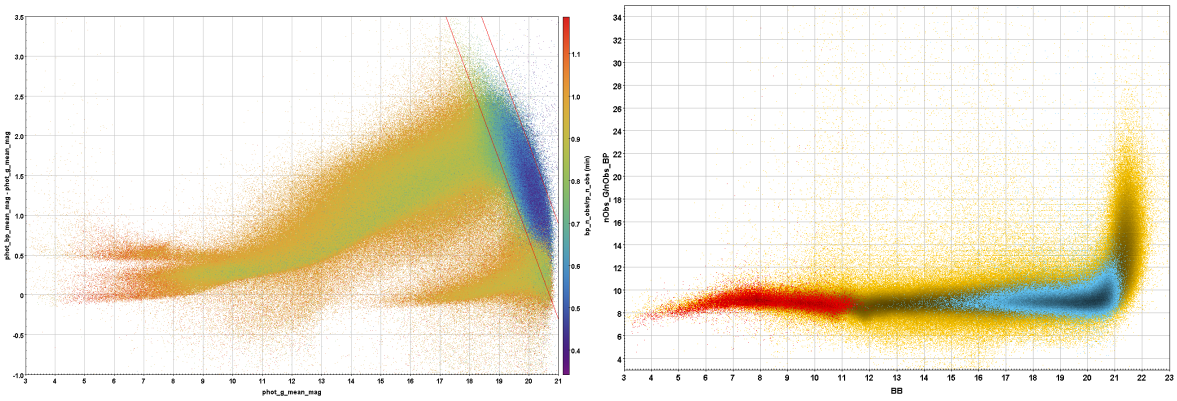

Figure 19: Left: Differences in magnitude between G and BP as a function of G, in EDR3 data, coloured by the minima in the ratio of BP over RP data points. Diagonal lines as in Fig. 13. Right: Ratio between observations in G and BP for the same transits. Colours as in Fig. 3.

The ”hook” as observed in Fig. 3 results from the data-rejection criteria used in Photpipe (see also Jira C5PPV-79). This can be seen in Fig. 19, where the differences between the G and BP magnitudes as a function of G magnitude (all in EDR3) are shown with respect to the ratio in observations for BP and RP. The hook is dominated by low numbers of BP compared to RP observations. The right-hand side of Fig. 19 shows how this reflects in the ratio of G over BP observations as a function of BP magnitude. Statistically this means that from the distribution of observed BP fluxes only those above a given limit have been used to construct the mean flux of the source. The epoch flux limit per sample used in Cycle 03 was 1.0 $e^{-}s^{-1}$. The hook starting shortly after the BP flux drops below 100 $e^{-}s^{-1}$ seems consistent with the rejection of integrated fluxes at epoch level at 1 $e^{-}s^{-1}$. By applying such rejection procedure the mean and uncertainty for the combined observations becomes biased by making the sources look brighter to much brighter (depending on the fraction of data rejected) than they really are, while at the same time reducing the uncertainty on the mean flux by restricting the range of flux values used. This feature is also discussed in Section 8.1 of (Riello et al., 2021).

The limit for the integrated flux appears to be close to a BP flux of 100, BP≈ 20.3 (signal-to-noise ratio for individual observations ≈ 2.6) in BP. In order to understand the relation between the rejection criterion and the appearance of the ‘hook’, the BP magnitudes for stars thus af-

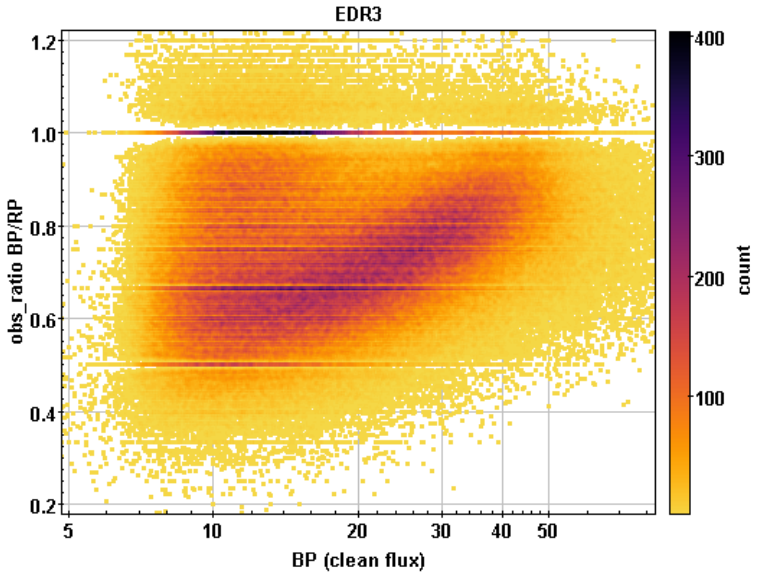

Figure 20: BP rejection levels as a function of BP flux, where the actual flux has been estimated based on the G flux.

fected were replaced by values based on an extrapolation of the relation between BP and G magnitudes:

$$\mathrm{BP}^* \approx 1.2\,\mathrm{G} - 1.95. \tag{22}$$

These artificial, clean, BP^* magnitudes were converted back to fluxes, and the flux values were compared with the rejection percentage, as based on a comparison between BP and RP observations. Figure 20 shows the results of the comparison: a clearly increasing rejection rate with decreasing real flux. The rejection rate as a function of real flux can be used to obtain an estimate of the actual mean noise level on these observations. That noise level is about a flux of 35 to 40, where observations below -1σ are rejected at a real flux of about 35 to 40, and 60 per cent of the data is rejected at a real count of 20. This noise level is most likely to be the result of the background contribution. If interpreted as a photon noise, it would imply an integrated background contribution of about 1225 to 1600, or per pixel 1.7 to 2.2. This is a reasonable mean background value for these data. A similar result was obtained in (Fabricius et al., 2021) through a data simulation, where a σ of 50 was used.

A comparison of the lower region of the HR diagram between stars with parallaxes larger than 10 mas and those with parallaxes between 4 and 5 mas clearly shows the effect the processing of the low BP fluxes has. What is still visible as a main sequence for the nearby stars has been lost in the ”hook” for the more distant stars (see Fig. 21). Excluding data with the BP over RP observations ratio less than 0.90 will largely remove the ”hook” from the HR diagram, but will also remove some good, valid data, for which the ratio reduction is just accidental.

Rejection of individual observations in the accumulation process can be done in such a way that it doesn’t bias the results, even for stars with low BP fluxes. The problem is in the assignment

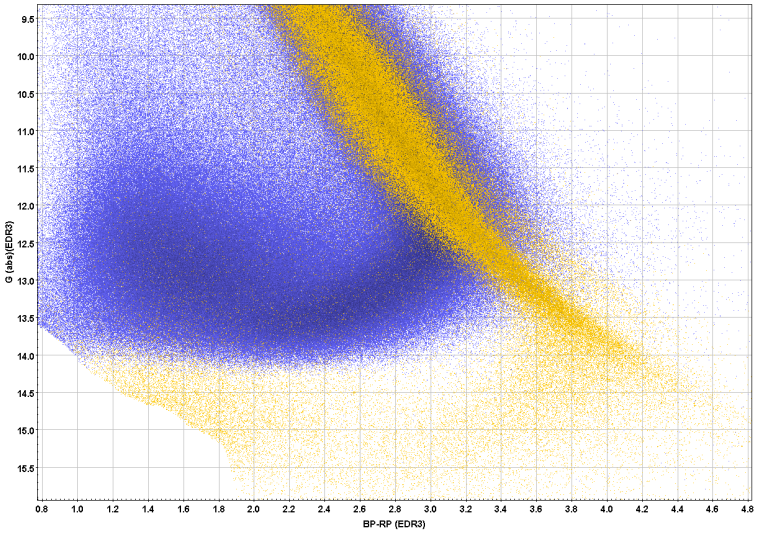

Figure 21: The lower part of the HR diagram for stars with parallaxes greater than 10 mas (ochre) and with parallaxes between 4 and 5 mas (blue). The lower main sequence shown by the nearby stars is lost for the more distant stars.

of errors to observations of low fluxes. By assigning an error based on the observed flux, the assumption is made that the observed flux is also the mean flux for that observation. For very faint fluxes this is not the case. To calculate a mean flux the assumption should be made that all observations stem from the same distribution, that for the mean flux of the source. The uncertainty on an individual flux measurement is then given by the mean flux plus the background contribution to the observation (background per sample times the number of samples used, see Section 4). To this can be added, if and when relevant, a calibration noise contribution, but that would normally be relatively small for very faint stars. A first estimate of the mean flux can be obtained by simply adding the observations without applying weights: the sum of all observed fluxes divided by the number of observations. In subsequent iterations the mean flux and background contributions can be used to assign a weight to each individual observation, to obtain a new mean flux estimate. Rejection of observations can take place at this stage, based on an evaluation of the weighted flux residual. The expected unit-weight standard deviation for this solution should get close to 1.0. Variable stars will produce an increased unit-weight standard deviation, but at the level of accuracy for individual transits, variability plays only a very minor role.

Figure 22 shows two more examples of how the data is affected by the ratio of BP over RP observations. For the flux excess factor, which compares BP, RP and G fluxes, there is no reference of unaffected data against which a comparison could be made to show that the values observed are actually indicating that the data is unreliable when this ratio of BP over RP observations gets below 0.9. The second graph shows how data with low BP over RP observation ratios is linked to the ”blob” in the distribution of dG against G. That would indicate that the ”blob” too is related to a problem with the BP magnitudes. In this case the situation is different:

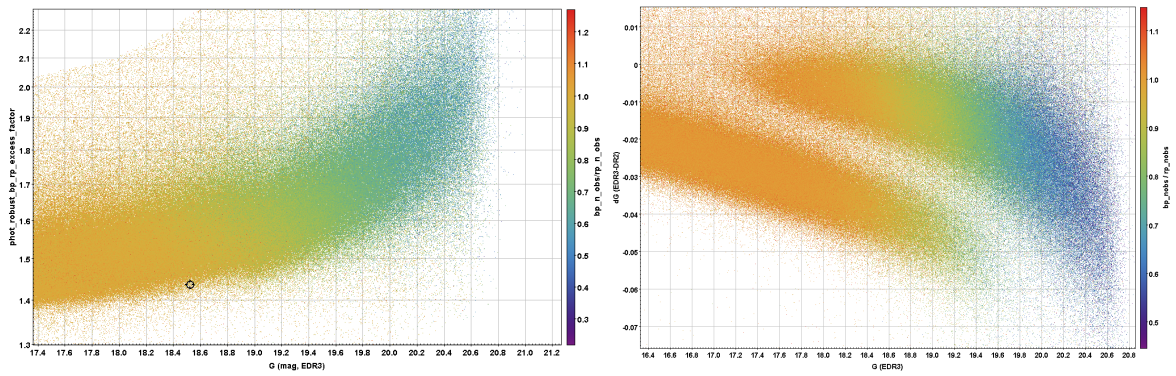

Figure 22: Left: The flux excess factor as affected by the ratio of BP over RP observations; Right: Similarly, for the blob in the relation between the BP and G magnitudes .

there is an almost fixed-level offset of about 0.025 magn between the ”blob” and what is likely to be unaffected data. That would typically represent a situation where the colour of the star as used in IPD was estimated, and the estimate was too blue. It should be noted here that no colour dependency was used in IDU for DR2, so this split is in the EDR3 data. It is the comparison with DR2 data that made this detection possible, as comparisons with other external catalogues tend to be much too noisy. From conversations with Michael Davidson it was learned that default colours assigned are indeed much less red than what should be applied for these very red stars. This would create a systematic discrepancy between the profile used for fitting the image and the actual signal of the image. That offset will be independent of brightness. An offset of 0.025 magn appears very feasible for such scenario. An *a posteriori* correction for this effect is presented in Riello et al. (2021).

10 Correcting, cleaning, filtering

Filtering based on the ratio of BP over RP observations would in principle be possible, but there is no clean cut border between the relative number of data affected by rejection based on the observed flux, and those rejected randomly for any other reason. Rejection also needs to consider the luminosity type of the star, since the data for White Dwarfs are affected differently from those of main sequence stars. Trials with different rejection criteria show either a significant residual ”hook” and/or the rejection of not-affected main sequence stars.

An a posteriori correction of the data is marginally possible, but would create rather noisy results. It could be obtained by interpreting the remaining data as obtained from the tail end of a Gaussian distribution of fluxes. By taking the RP observations as reference, the number of observations lost can be estimated. Assuming the sigma of the fluxes to be related to the mean flux and average background, as shown above, would provide the possibility to reconstruct the underlying flux. Such estimate is likely to be disturbed by fluctuating background contributions and the estimate of the number of true observations.

It is important that this issue is corrected before the DR3 data are used as input for IDU in DR4, as otherwise the lack of reliable BP magnitudes and BP−RP colours for faint red stars will still affect aspects of the DR4 release. The recommendation would be to correct the BP fluxes of faint red stars for DR3, replacing those in EDR3. A small number of RP fluxes for faint blue stars may also be affected and in need of replacement, and any flux accumulation that implements the systematic rejection of low fluxes should be re-assessed.

11 Conclusions

The comparison between the DR2 and EDR3 photometry showed a discontinuity in the G magnitudes caused by the use of default colours in IPD. As no colour dependencies were used in IPD for DR2, this has to be in the EDR3 data. If the default colour used in IPD is preserved in the data record, it is possible to correct for this kind of discontinuous behaviour.

In EDR3, as well as DR2, there is an important issue, to do with faint red stars, present in the broad-band photometric data, which is possible to resolve in future data releases. The current approach in the accumulation of the BP fluxes for faint red stars, creates significant bias in the mean flux estimates and colour indices for these stars. A correction afterwards is very likely to be too noisy to be of use. Filtering out only the BP data leaves only the much less powerful G−RP as a colour index, which may still be useful in some applications, but which should only be seen as a fallback. A procedure for accumulating low flux data is described in the text. Another issue is directly related to this. With unreliable colour indices for faint red stars, default colours will instead be used in IPD, and those default colours are often significantly different from the actual colours of the stars, creating underestimated fluxes in the G band. This, as was stated above, was noted in the comparison with the DR2 data, which had no colour dependency in IPD. This situation is not completely avoidable, but can be dealt with by preserving information on the colour as used in IPD. A correction to the mean flux afterwards can then in principle be made once a reliable colour for the star has been derived.

Finally, some of the issues noted in the data are a reflection of the high quality of the Gaia data and may not be seen when comparing the Gaia data with external data. The high accuracy exposes issues that usually remain hidden behind the noise.

Acronyms

The following is a complete list of acronyms used in this document. Use the acronym tool from CU1 to generate acronym list for your doc. The following table has been generated from the on-line Gaia acronym list:

Acronym	Description
AC	ACross scan
ADU	Analogue-to-Digital Unit
BP	Blue Photometer
CCD	Charge-Coupled Device
DR2	Gaia Data Release 2
DR3	Gaia Data Release 3
DR4	Gaia Data Release 4
EDR3	Gaia Early Data Release 3
ELSF	Empirical LSF (”option 4”)
G	G-band photometry
HR	High Resolution
HRD	Hertzsprung-Russell Diagram
IDU	Intermediate Data Update
IPD	Image Parameter Determination
LMC	Large Magellanic Cloud (special, high-density area on the sky)
OBMT	On-Board Mission Timeline
PSF	Point Spread Function
PhotPipe	Photometric Pipeline (CU5 / DPCI)
RGB	Red Giant Branch (star)
RP	Red Photometer
SMC	Small Magellanic Cloud (special, high-density area on the sky)
SPSS	Spectro-Photometric Standard Stars
TN	Technical Note

References

Casagrande L., VandenBerg D.A., 2018, MNRAS, 479, L102, ADS Link

Fabricius C., Luri X., Arenou F., et al., 2021, A&A, 649, A5, ADS Link

Jordi C., Fabricius C., Figueras F., et al., *Error model for photometry and spectrophotometry*, GAIA-C5-TN-UB-CJ-047, URL http://www.rssd.esa.int/cs/livelink/open/2871236

van Leeuwen F., *The effects of gate deployment on image saturation during data segment 3*, GAIA-C5-TN-IOA-FVL-181, URL http://www.rssd.esa.int/cs/livelink/open/1387551

Lindegren L., Bastian U., Biermann M., et al., 2021, A&A, 649, A4, ADS Link

Riello M., De Angeli F., Evans D.W., et al., 2021, A&A, 649, A3, ADS Link

Rowell N., Davidson M., Lindegren L., et al., 2021, A&A, 649, A11, ADS Link