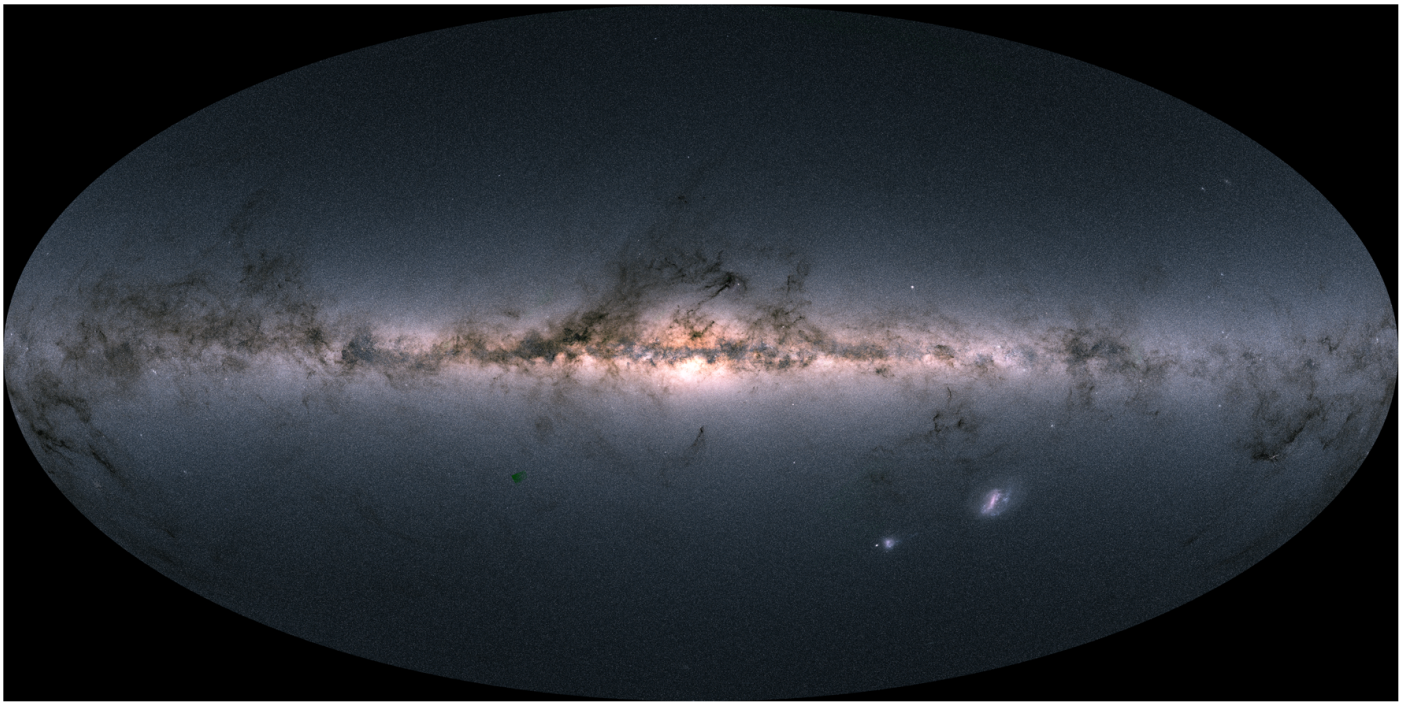

Fig. 4. Map of the total flux measured in the G_{RP}, G, and G_{BP} bands, where the flux in these bands is encoded in the red, green, and blue channel, respectively. There is one easily visible artefact in this map, a "green" patch to the lower left of the bulge which is a region where G_{BP} and G_{RP} data are not available for a large number of sources, leading to the greenish colour which was used to encode the G-band fluxes (which are available for all sources). Such artefacts also occur (although not as visible) in the region to the upper left of the Small Magellanic Cloud and at high Galactic latitude to the right of the north Galactic pole region. The areas where green patches are likely to occur can be identified in Fig. 27 in Evans et al. (2018) which shows the celestial distribution of *Gaia* DR2 sources for which no BP/RP photometry is available.

Figure 4 shows a map that combines the integrated fluxes as observed in the G_{RP}, G, and G_{BP} bands, where the integrated flux map for each of the bands was used to colour code the image according to a red, green, and blue channel. The map illustrates the availability of homogeneous all-sky multi-band photometry in *Gaia* DR2 and offers a magnificent view of the Milky Way in colour. This flux map also reveals numerous open clusters which are not readily visible in the source count map (while on the other hand many faint source concentrations, such as distant dwarf galaxies are no longer visible). Complete details on the construction of the images in Figs. 3 and 4 are provided in Moitinho et al. (2018).

One aspect of the sky maps shown in Figs. 3 and 4 that is perhaps not as well appreciated is their effective angular resolution, which given the size of *Gaia*'s main telescope mirrors (1.45 m along the scanning direction, Gaia Collaboration 2016b) should be comparable to that of the *Hubble* Space Telescope. Gaia Collaboration (2016a) and Arenou et al. (2017) discuss how the effective angular resolution of *Gaia* DR1 is limited to about 2–4 arcsec owing to limitations in the data processing. This has much improved for *Gaia* DR2. The gain in angular resolution is illustrated in Fig. 5. The top panel shows the distribution of source pair distances in a small, dense field. For *Gaia* DR2 (upper, red curve) source pairs below 0.4–0.5 arcsec are rarely resolved, but the resolution improves rapidly and above 2.2 arcsec practically all pairs are resolved. For *Gaia* DR1 the fraction of resolved source pairs started to fail at separations of 3.5 arcsec, reaching very low values below 2.0 arcsec. The same, modest resolution is seen for *Gaia* DR2 if we only consider sources with G_{BP} and G_{RP} photometry. The reason is the angular extent of the prism spectra and the fact that *Gaia* DR1 only includes sources for which the integrated flux from the BP/RP spectra could be reliably determined. The lower panel shows in the same way the source pairs in the one hundred times larger, sparse field. The more remarkable feature here is the peak of resolved binaries at small separations, which was missed in *Gaia* DR1. A similar population must be present in the dense field, where it cannot be discerned because the field is dominated by distant sources. The figure also demonstrates that the gain in number of sources from *Gaia* DR1 to *Gaia* DR2 is mainly due to the close source pairs. Finally, Fig. 5 clearly demonstrates that the effective angular resolution of *Gaia* DR2 quite significantly exceeds that of all ground-based large-area optical sky surveys.

5. Treat *Gaia* DR2 as independent from *Gaia* DR1

Although *Gaia* DR1 and *Gaia* DR2 are based on observations from the same instruments, the discussion in the following subsections shows that the two releases should be treated as independent. In particular the tracing of sources from *Gaia* DR1 to *Gaia* DR2 (should this be needed for a particular application) must be done with care.

5.1. Gaia DR2 represents a stand-alone astrometric catalogue

Because the observational time baseline for *Gaia* DR2 is sufficiently long, parallax and proper motion can be derived from the *Gaia* observations alone. That is, the *Tycho-Gaia* Astrometric Solution (TGAS, Michalik et al. 2015a) as employed for the 2 million brightest stars in *Gaia* DR1 is no longer needed, and the astrometric results reported in *Gaia* DR2 are based solely on *Gaia* observations. For the TGAS subset from *Gaia* DR1 there is thus a large difference in the time baseline for the proper motions

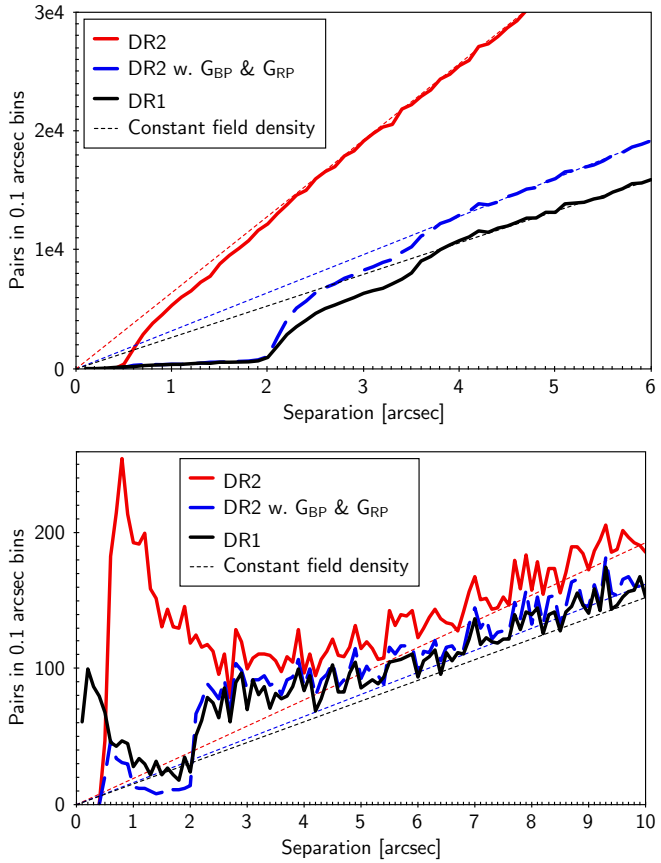

Fig. 5. Histograms from Arenou et al. (2018) of source pair separations in two circular test fields for *Gaia* DR2 sources (red lines); *Gaia* DR2 sources with G_{BP} and G_{RP} photometry (blue lines); and *Gaia* DR1 sources (black lines). *Top*: a dense field of radius 0.5° at $(\ell, b) = (-30°, -4°)$ with 456 142 sources, *bottom*: a sparse field of radius 5° at $(\ell, b) = (-100°, -60°)$ with 250 092 sources. The thin, dotted lines show the relations for a constant density across the field.

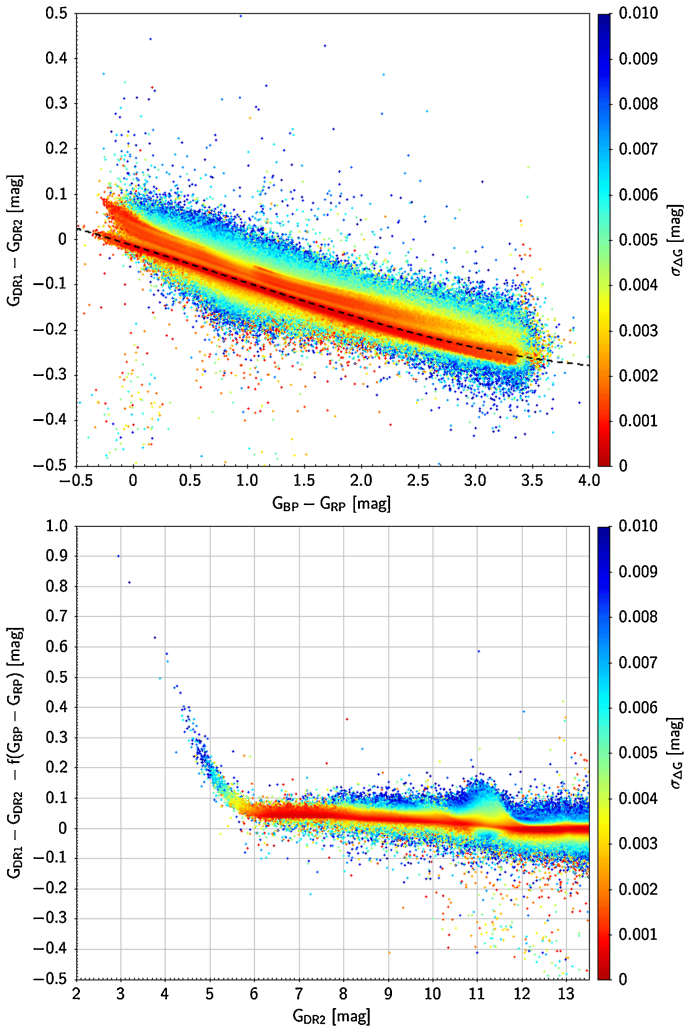

Fig. 6. *Top diagram*: difference in the value of G (with $\Delta G = G_{DR1} - G_{DR2}$) as listed for the same sources in *Gaia* DR1 and *Gaia* DR2 as a function of $(G_{BP} - G_{RP})$. The source pairs selected from the two releases match in celestial position to within 0.25 arcsec and the formal error on the magnitude differences is less than 0.01. All sources were selected to have a flux excess factor in *Gaia* DR2 of less than 1.6. The dashed line shows a polynomial relation between the difference in G and the colour. The colour scale indicates the estimated uncertainty on ΔG. *Bottom panel*: relation between ΔG and G after removing the colour dependency using the polynomial relation in the top panel.

(~24 yr vs ~2 yr) which means there can be significant differences between TGAS and *Gaia* DR2 proper motions for binary stars with orbital periods comparable to 2 yr. The TGAS proper motions may be more reliable in such cases. However, discrepancies can also point to erroneous TGAS proper motions related to a mismatching between (components of) sources observed by *Gaia* and HIPPARCOS (see Makarov et al. 2017, for a discussion of this issue). In cases where proper motion discrepancies are of interest they should be carefully investigated before deciding which values to use or concluding that the discrepancy points to the source not being a single star.

5.2. Photometric system evolution

The photometric data processing for *Gaia* DR2 (Riello et al. 2018; Evans et al. 2018) features many improvements with respect to *Gaia* DR1 and represents a new photometric reduction. In particular more input data was used and the stretch of data selected for the initialisation of the photometric calibration was largely free of the effects of contamination by water ice (see Gaia Collaboration 2016b, for a summary of the contamination problem in the early phases of the *Gaia* mission). As a consequence the photometric system for *Gaia* DR2 is different from *Gaia* DR1. This is illustrated in Fig. 6 which shows the difference in G-band magnitude ($\Delta G = G_{DR1} - G_{DR2}$) for the same sources between the two data releases. The source pairs selected from the two releases match in celestial position to within 0.25 arcsec and the formal error on the magnitude differences is less than 0.01. All sources were selected to have a flux excess factor in *Gaia* DR2 of less than 1.6 (see Sect. 3.2 for a description of this quantity). The two panels in Fig. 6 show that there is a substantial difference in the G band values, with the mean of ΔG being about −0.1 mag, and a strong colour dependence which is indicated by the dashed line showing the polynomial relation

$$\begin{aligned} G_{DR1} - G_{DR2} = &-0.013612 - 0.079627(G_{BP} - G_{RP}) \\ &- 0.0040444(G_{BP} - G_{RP})^2 \\ &+ 0.0018602(G_{BP} - G_{RP})^3 . \end{aligned} \quad (4)$$

Removing the colour dependence and plotting ΔG vs. G (bottom panel of Fig. 6) reveals image saturation effects at the bright end which more strongly affect the *Gaia* DR1 magnitudes. Sources

with larger magnitude differences typically have large estimated uncertainties (blue points) whereas the majority of sources have smaller differences and small estimated errors (red points). The feature near $G \approx 11.5$ mag is due to the high and variable photometric uncertainties in *Gaia* DR1 for bright sources (see Fig. 9 in Evans et al. 2018).

This difference in photometric systems means that one should not apply photometric calibrations derived from *Gaia* DR1 (e.g., the calibration of the red-clump absolute G-band magnitude, Ruiz-Dern et al. 2018; Hawkins et al. 2017) to *Gaia* DR2 photometry. The G passband calibration also changes from *Gaia* DR1 to *Gaia* DR2. The passbands for G, G_{BP}, and G_{RP} are described in Evans et al. (2018). They are available in the version that was used for the *Gaia* DR2 data processing and in a revised version which was determined after the processing was finished (see Sect. 6.3.2). The revised passband should be used for precise photometric work based on the fluxes listed in *Gaia* DR2. The nominal (pre-launch) passband as provided on the *Gaia* science performance pages[2] and independent passband calibrations based on *Gaia* DR1 (Weiler et al. 2018; Maíz Apellániz 2017) should not be used. Likewise the nominal transformations between the *Gaia* broad-band photometry and other photometric systems listed in Jordi et al. (2010) should not be used. Refer to Evans et al. (2018) for the updated relations. To take full advantage of the high precision *Gaia* DR2 photometry, predictions of the *Gaia* broad-band magnitudes for stellar evolutionary tracks or isochrones in the colour-magnitude diagram (e.g. Choi et al. 2016; Marigo et al. 2017) should be updated.

5.3. Source list evolution

The processing for a given data release starts with a task that groups individual *Gaia* observations and links them to sources on the sky (see Lindegren et al. 2018; Fabricius et al. 2016, for a description of this process). The observations are linked to known sources, or sources are newly "created" from the clustering of the observations around a celestial position where previously no source was known to exist. This leads to a working catalogue of sources (hereafter called "the source list") and their corresponding observations, which forms the basis for the subsequent data processing. In this list the sources are assigned a *Gaia* source identifier which is intended to be stable for every source. The algorithm that carries out the grouping and linking was much improved at the beginning of the processing for *Gaia* DR2. The improved source list will lead to the following changes in linking the observations to the source identifiers for a substantial fraction of sources:

- The merging of groups of observations previously linked to more than one source will lead to a new source associated to the merged observations (with a new source identifier) and the disappearance of the original sources (along with their source identifiers).
- The splitting of groups of observations previously linked to one source will lead to new sources associated to the split groups of observations (with new source identifiers) and the disappearance of the original source (along with its source identifier).
- The list of observations linked to a source may change (and hence the source characteristics may change), while the source identifier remains the same.

In the processing for *Gaia* DR2 the number of changes of source identifiers (where the physical source remains the same) is large. At magnitudes brighter than $G \approx 16$ some 80–90 per cent of the sources changed source identifier. At $G \approx 18$ mag this reduces to some 20 per cent, going down to zero source identifier changes around $G = 20$ mag.

The consequence is that one should not blindly use the source identifier to look up sources from *Gaia* DR1 in *Gaia* DR2. Example applications we have in mind are the repeat of an analysis done with the first data release using the new data and the retrieval of a list *Gaia* DR1 sources, cross-matched against some other survey, from the *Gaia* DR2 tables. The recommendation is to treat the source lists from the two releases as completely independent. An additional field will be added to *Gaia* DR2 and subsequent releases which specifies the *Gaia source name* as "*Gaia* DR*n* `source_id`". The bare source identifier can be used for efficient queries of the large *Gaia* data base, while the source name should always be specified (i.e., including the data release number) when referring to the source in the literature. To facilitate the tracing of sources from *Gaia* DR1 to *Gaia* DR2 a table is provided which lists for each *Gaia* DR2 source the potential matching sources in *Gaia* DR1 (and vice versa). For the majority of sources (over 99 per cent) there is a one-to-one correspondence (although the source identifier can differ), but multiple matches may occur and then it is up to the user of the *Gaia* data to make a judgement as to which pair is the correct match (where the possible differences in the G-band magnitude should be kept in mind).

The source list is expected to stabilise in future *Gaia* data releases with much less change expected between *Gaia* DR2 and *Gaia* DR3. However some evolution of the source lists will take place up to the final data release and we stress that a change in source *character* can always occur as observations are added in future data releases (e.g., a stable source can turn into a variable from one data release to the next).

6. Using *Gaia* DR2 data: completeness and limitations

Gaia DR2 represents a major advance compared to *Gaia* DR1, featuring new data types and a much expanded and improved astrometric and photometric data set. Nevertheless this release is still intermediate, based on only a limited amount (~22 months) of input data, and still suffers from simplifications in the data processing that will introduce shortcomings in the calibrations which in turn can introduce systematic errors. We summarise here the main limitations of *Gaia* DR2 which the user of the data should be aware of.

6.1. Gaia DR2 validation and source filtering

The validation of the *Gaia* DR2 results followed the process described in Gaia Collaboration (2016a). We refer to the papers listed in Sects. 2 and 3 for full details on the validation of the data done at the level of the individual data processing systems. The overall validation, assessing the combined results is described in Arenou et al. (2018). As was the case for *Gaia* DR1 the results validation revealed no problems that prevented a timely release of *Gaia* DR2, but filtering of the available data processing outputs before their incorporation into *Gaia* DR2 was still necessary. The level of filtering is significantly reduced compared to that for *Gaia* DR1 as can be appreciated from the substantial increase in the number of sources for which astrometric and photometric data is published. We summarise the filtering that was

[2] https://www.cosmos.esa.int/web/gaia/science-performance

applied with the aim of providing a better understanding of some of the survey characteristics.

6.1.1. Astrometry

For the astrometric data set the results were filtered by requiring that a source was observed by *Gaia* at least five times (five focal plane transits), and that the astrometric excess noise and the semi-major axis of the position uncertainty ellipse are less than 20 and 100 mas, respectively. In addition within the astrometric solution pipeline the parallax and proper motions are determined only for sources satisfying the requirement that they are brighter than $G = 21$, that the number of "visibility periods" used is at least 6 (a visibility period represents a group of observations separated from other such groups by at least four days), and that the semi-major axis of the 5-dimensional uncertainty ellipse is below a magnitude dependent threshold. We refer to Lindegren et al. (2018) for the details. For sources that do not meet these requirements only the positions are reported in *Gaia* DR2.

6.1.2. Photometry

The photometric inputs were filtered as follows. Sources without a well-determined value for G do not appear in *Gaia* DR2. The photometry in the G, G_{BP}, or G_{RP} bands is only reported if the source was observed at least twice by *Gaia* in the respective bands. For the so-called "bronze" sources (see Sect. 3.2 and Riello et al. 2018) no colour information (i.e. no G_{BP} and G_{RP}) is reported. This also holds for sources fainter than $G = 21$ mag and sources for which the flux excess factor is above 5. Hence *Gaia* DR2 contains a substantial number of sources (~300 million) for which no colour information is available. Note however that the filtering on flux excess factor was not applied to the variable source time series tables, hence there may be sources that have no G_{BP} and/or G_{RP} value listed but for which a light curve in G_{BP} and/or G_{RP} is nevertheless reported.

6.1.3. Radial velocities

For sources satisfying the following conditions no radial velocity is reported in *Gaia* DR2. The source is fainter than $G_{RVS} = 14$ (the limit refers to the flux as actually measured in the RVS band, not the provisional G_{RVS} value mentioned in Sect. 3.3); the fraction of transits where the source was detected as having a double-lined spectrum was larger than 0.1 (this removes detected double-lined spectroscopic binaries); the uncertainty on the radial velocity is above 20 km s^{-1}; the effective temperature corresponding to the spectral template used to derive the radial velocity is outside the range 3550–6900 K. By construction the RVS data processing limited the range of possible radial velocities to $|v_{rad}| < 1000$ km s^{-1}. Special care was taken for the 613 sources that had measured radial velocities with absolute values above 500 km s^{-1}. Because this small subset can easily be contaminated by outliers caused by data processing limitations, their spectra were visually inspected. Of these 613 sources, 202 were included in *Gaia* DR2 as valid high velocity sources, while the remainder were removed from the published catalogue. For sources with radial velocities at absolute values below 500 km s^{-1} visual inspection was not possible due to the progressively (much) higher numbers. The users of *Gaia* DR2 should thus be aware of the specific selection applied to sources with $|v_{rad}| > 500$ km s^{-1}. We refer to Katz et al. (2018) for more details on this issue.

6.1.4. Variable stars

During the variability analysis a strict internal filtering was applied to the quality of the photometric time series (such as removing negative or unrealistically low flux values). This was followed by a filtering of the classification results to reduce the contamination due to data processing artefacts and confusion between variable types. The outputs from the specialised variable star characterisation pipelines were filtered to remove sources for which the results of the light curve analysis were not deemed reliable enough. This combination of filters reduced the number of sources flagged as variable to the numbers listed in Table 1. The reader interested in using the variable star data set is strongly advised to consult Holl et al. (2018) and references therein, as well as the online documentation.

6.1.5. Astrophysical parameters

The astrophysical parameter results are only presented for sources brighter than $G = 17$ (no fainter sources were processed) and only for sources for which G, G_{BP}, and G_{RP} are reported. Further filtering was applied based on the quality of the various inputs to the astrophysical parameter estimation, where particularly strict criteria were applied to the extinction and reddening estimations. The details of the filtering applied to the astrophysical parameters are best understood in conjunction with the description of how these parameters were estimated. Hence we refer to Andrae et al. (2018) for the details (see also Sect. 6.3.4).

6.1.6. Solar system objects

For the solar system data set the filtering on input data quality (internal to the processing pipeline) was followed only by the removal of some SSO observations for which the relative flux uncertainty in the G band was larger than 0.1. This mainly removes observations of the very "fast" SSOs for which the observation window may be badly placed (causing flux loss) toward the end of the focal plane transit. In addition a selection of the SSO observations was removed as well as some individual sources (see Gaia Collaboration 2018f, for details).

6.1.7. Duplicated sources

A global filter concerns the removal of duplicates of sources, which sometimes occur when the observation to source matching process creates two clusters of detections that later turn out to belong to the same source (see Gaia Collaboration 2016a; Fabricius et al. 2016). The 47 802 437 sources for which the duplicate was removed are indicated as such. The removal of duplicates is done after the completion of the data processing. Hence the observations corresponding to the removed component are effectively not used for, and do not appear in, the published catalogue. In future *Gaia* data releases the duplicates are expected to be merged into a single source.

6.2. Survey completeness

As can be appreciated from Fig. 1 the completeness of the *Gaia* survey has much improved for the second data release, being essentially complete between $G = 12$ and $G = 17$. The completeness at the bright end has improved, although a fraction of the bright stars at $G < 7$ is still missing with no stars brighter than $G = 1.7$ mag appearing in *Gaia* DR2. Gaia Collaboration (2016a) extensively explain how the combination of the *Gaia*

scan law coverage of the sky over the period covered by *Gaia* DR1 combined with the filtering applied to the astrometric and photometric results leads to strips or holes with a lack of sources (see Fig. 11 and 12 in that paper). Although much reduced (as seen in Fig. 3), these artefacts are still present in the *Gaia* DR2 source list and start appearing at $G > 17$.

We list here a number of more specific remarks on the completeness of *Gaia* DR2:

- The completeness for high proper motion stars has significantly improved with respect to *Gaia* DR1, but it is estimated that some 17 per cent of high proper motion stars (with $\mu > 0.6$ arcsec yr^{-1}) are still missing (for various reasons).
- In crowded regions the capability to observe all stars is reduced (Gaia Collaboration 2016b). In combination with the still limited data treatment in crowded areas (see section 6.2 in Gaia Collaboration 2016a) this means that the survey limit in regions with densities above a few hundred thousand stars per square degree can be as bright as $G = 18$.
- As described in Sect. 4 the effective angular resolution of the *Gaia* DR2 source list has improved to ∼0.4 arcsec, with incompleteness in close pairs of stars starting below about 2 arcsec. Refer to Arenou et al. (2018) for details.
- We repeat that the radial velocity, astrophysical parameter and variable star data sets are far from complete with respect to the overall *Gaia* DR2 catalogue (see Sect. 3 above). In particular the radial velocities are only reported for a restricted range in effective temperatures (of the spectral templates, see Sect. 6.1.3) and the completeness of the radial velocity catalogue with respect to *Gaia* DR2 varies from 60 to 80 per cent (Katz et al. 2018) over the range $G = 4$ to $G = 12$.
- The solar system object sample processed for *Gaia* DR2 was pre-selected and is not a complete sample with respect to criteria like dynamics, type, category, etc. In addition bright SSOs ($G \lesssim 10$) were removed from the published results because the astrometry in that brightness range is limited in quality by calibration uncertainties and systematics related to the apparent source size and motion on the sky (leading to the use of inadequate PSF models for the image centroiding).

For more detailed information on the completeness of *Gaia* DR2 we refer to the individual data processing papers and the overall validation paper (Arenou et al. 2018). No attempt was made at deriving a detailed survey selection function.

6.3. Limitations

6.3.1. Astrometry

The astrometry in *Gaia* DR2 represents a major improvement over *Gaia* DR1 with an order of magnitude improvement in the uncertainties at the bright end and a vast expansion of available parallaxes and proper motions. In particular the individual uncertainties are much closer to having been drawn from Gaussian distributions and the systematics in the parallax uncertainties are now generally below the 0.1 mas level (as estimated from the analysis of QSO parallaxes, Lindegren et al. 2018). However, the users of the *Gaia* DR2 astrometry should be aware of the following. There is an overall parallax zeropoint of ∼−0.03 mas (as estimated from QSO parallaxes, in the sense of the *Gaia* DR2 parallaxes being too small) which the data have not been corrected for (see below), and the astrometry shows systematics correlated to celestial position, source colour, and source magnitude. Moreover the parallaxes and proper motions show significant spatial (i.e. source-to-source) correlations of up to 0.04 mas and 0.07 mas yr^{-1} over angular scales from <1 to 20 degrees (see Lindegren et al. 2018, for a more detailed characterisation of the spatial covariances). These regional systematics are visible in maps of average QSO parallaxes and in dense fields where the large amount of sources allows to average the astrometric parameters and visualise the systematic differences in, for example, the parallax zeropoint (Lindegren et al. 2018; Arenou et al. 2018).

One might expect that the published parallax values would have been adjusted according to the global zeropoint, however a deliberate choice was made not to apply any corrections to the *Gaia* DR2 astrometry. This is motivated by the fact that the value of the zeropoint depends on the sample used to estimate its value (Arenou et al. 2018). The differences are related to the dependence of the systematics in the astrometry on source position, colour, and magnitude, meaning that the zeropoint for QSOs (faint, blue) may not be representative of the zeropoint for a sample of bright red stars. In addition the correction of the global zeropoint would represent an arbitrary choice with respect to the regional systematics which would be left uncorrected.

The astrometric uncertainties listed in *Gaia* DR2 are derived from the formal uncertainties resulting from the astrometric data treatment, and unlike for *Gaia* DR1 these have not been externally calibrated (by comparison to the HIPPARCOS data, Lindegren et al. 2016). At a late stage during the preparation of *Gaia* DR2 a bug was discovered in the astrometric processing software. This did not significantly affect the astrometric parameters themselves but resulted in a serious underestimation of the uncertainties for the bright sources ($G \lesssim 13$). Rather than recomputing the full solution, with serious repercussions for the downstream processing and publication schedule, it was decided to apply an approximate ad hoc correction to the uncertainties. The details of this are described in Appendix A of Lindegren et al. (2018). While the corrected (published) uncertainties are thus approximately consistent with the residuals of the astrometric solution, comparisons with external data show that they are still underestimated (Arenou et al. 2018). The underestimation is moderate (∼7–10%) for faint sources ($G > 16$) outside the Galactic plane, but may reach 30 to 50 per cent for sources of intermediate magnitude ($G \simeq 12$–15). At brighter magnitudes a comparison with HIPPARCOS data suggests that uncertainties are underestimated by no more than 25 per cent (Arenou et al. 2018). No additional correction was made in the published data based on these external comparisons, and users of the data may have to allow for it in their analyses.

The PSF model used in the pre-processing is essentially the same as that used for *Gaia* DR1, and the iterative loop between the astrometric and photometric data treatment and the pre-processing is not yet closed (see Sect. 6.1 and Fig. 10 in Gaia Collaboration 2016a). This implies that the PSF calibrations and the subsequent determination of the source flux and location have not benefited from better input astrometry and source colours. These inadequacies in the instrument calibration have a particularly large impact on the astrometry of bright stars ($G \lesssim 13$) which is visible in the uncertainties being larger than those for somewhat fainter stars. In addition there may be a systematic rotation of the proper motion system for the bright stars with respect to QSOs (see Table 3 and Lindegren et al. 2018), and the parallax zeropoint may be different.

6.3.2. Photometry

The strongly varying photometric uncertainty at the bright end in G and the bumps in the uncertainty around $G \sim 13$ and

$G \sim 16$ visible in the *Gaia* DR1 data (Gaia Collaboration 2016a; Evans et al. 2017) are still present although in much reduced form (Evans et al. 2018). The uncertainties on G_{BP} and G_{RP} as a function of magnitude are much smoother with the integrated prism photometry being much less sensitive to instrument configuration changes.

The flux excess factor can take extreme values and it was decided not to publish colour information for sources with a flux excess factor above 5 (this is a rather liberal filtering). We recommend that the value of the flux excess factor is used to clean samples of sources selected from *Gaia* DR2 from the most problematic cases, in particular if accurate colour information is important. The flux excess factor has a dependence on $(G_{BP} - G_{RP})$, which any filtering should take into account. We refer to Evans et al. (2018), Gaia Collaboration (2018a), and Lindegren et al. (2018) for more detailed recommendations on cleaning samples from the effects of the flux excess in the BP/RP bands.

Although not really a limitation in the photometric data, we nevertheless point out the following in relation to the photometric zeropoints and passbands. The photometric zeropoints used to convert the photometric fluxes into the magnitudes listed in *Gaia* DR2 are derived from the pass bands used internal to the processing for this release. The calibration of these passbands was done in a preliminary manner and they have been updated after the *Gaia* DR2 processing was completed through an analysis employing BP/RP spectra which were not available for the earlier calibrations.The magnitude zeropoints for the updated passbands differ by up to 3 mmag from those used to calculate the *Gaia* DR2 magnitudes (Evans et al. 2018). As remarked in Sect. 5, for precision photometric work the updated passbands should be used and then the difference in zeropoints should be accounted for (by recalculating the magnitudes from the fluxes listed in *Gaia* DR2).

6.3.3. Radial velocity data

When using the radial velocities from *Gaia* DR2 the following limitations should be taken into account. Single-lined spectroscopic binaries have been treated as single stars and only the median radial velocity, together with information on the scatter in the underlying (but unpublished) epoch radial velocities, is provided. Double lined spectroscopic binaries which were detected as such were not processed and are missing from the *Gaia* DR2 radial velocity data set. Double lined spectroscopic binaries with a weak secondary component are present in the catalogue and have also been treated as single stars. No radial velocities have been determined for stars with detected emission lines and there are no radial velocities for "cool" and "hot" stars (Sect. 6.1.3). Radial velocities with absolute values above 500 km s^{-1} should be treated with some care. Beyond this limit clearly dubious values were filtered out of the catalogue but it is not guaranteed that all remaining radial velocities above +500 km s^{-1} or below −500 km s^{-1} are reliable.

Through comparison with other radial velocity surveys it is concluded that the *Gaia* DR2 radial velocities are accurate to a few 100 m s^{-1}, where systematic differences can be due to both *Gaia* DR2 and the other surveys. Katz et al. (2018) show that while offsets are lower than 300 m s^{-1} for bright stars ($G_{RVS} < 10$), a trend with magnitude is seen in all the comparisons with other surveys, reaching ∼500 m s^{-1} at the faint end.

Finally, we note that *Gaia* DR2 lists the atmospheric parameters (T_{eff}, $\log g$, [Fe/H]) of the spectral templates used in the derivation of the radial velocities through the cross-correlation technique. Their values should *not* be used as estimates of the actual atmospheric parameters of the stars, they are only provided as extra information to judge the quality of the radial velocities.

6.3.4. Astrophysical parameters

The values of T_{eff}, A_G, $E(G_{BP} - G_{RP})$, radius, and luminosity were determined only from the three broad-band photometric measurements and the parallax, on a star-by-star basis (where parallax was not used to estimate T_{eff}). The strong degeneracy between T_{eff} and extinction/reddening when using the broad band photometry necessitates rather extreme assumptions in order to estimate their values. This can lead to correspondingly strong systematics in the astrophysical parameters which are not accounted for in the uncertainties listed in *Gaia* DR2. We summarise here the most important caveats but refer to the online documentation and Andrae et al. (2018) for more extensive guidelines on the use of the astrophysical parameter estimates. The assessment of the quality of the astrophysical parameters from the perspective of the overall validation of *Gaia* DR2 can be found in Arenou et al. (2018).

The estimation of T_{eff}, A_G, and $E(G_{BP} - G_{RP})$ was done using a machine learning algorithm (specifically, the extremely randomised trees, or ExtraTrees algorithm Geurts et al. 2006). For the T_{eff} estimation the algorithm was trained on the photometry for *Gaia* sources for which T_{eff} estimates were available from existing independent surveys (see Andrae et al. 2018, Table 2). Only effective temperatures over the range 3000–10 000 K were considered and the training data shows strong peaks at specific T_{eff} values. The training set for the extinction and reddening estimation was based on synthetic photometry constructed using PARSEC 1.2S[3] stellar models which are accompanied by simulated photometry based on the Atlas 9 synthetic spectral library (Castelli & Kurucz 2003). No attempt was made at a realistic population of the synthetic colour magnitude diagrams in terms of the stellar initial mass function, the metallicity distribution, or the frequency of extinction values. All sources were treated as single stars and no attempt was made to filter out known galaxies, binaries, etc. Please refer to Andrae et al. (2018) for full details.

No T_{eff} values outside the range 3000–10 000 K are reported as these were not contained in the training data used for the estimation algorithm. This means that stars with effective temperatures outside the aforementioned range will have systematically too high or too low T_{eff} values listed in *Gaia* DR2. The distribution of T_{eff} values contains artefacts that reflect the distribution of the T_{eff} values in the training data. Effective temperature estimates in high extinction areas can be underestimated as the training data contained no examples of extincted stars.

The estimates of A_G and $E(G_{BP} - G_{RP})$ have such large uncertainties in general that their usefulness for individual stars is very limited. The extinction/reddening estimates should be used statistically only (for collections of stars) in which case the extinction maps shown in Andrae et al. (2018) demonstrate that on average the A_G estimates are reliable. The extinction estimates are strictly non-negative (with a model grid imposed maximum of $A_G = 4$) and have non-Gaussian posteriors, for which asymmetric uncertainties are listed in the catalogue. The non-negativity constraint can lead to apparent overestimation of the extinctions in regions, such as at high Galactic latitudes, where low extinction is expected on average. The effective temperature and extinction signals are degenerate in the broadband

[3] http://stev.oapd.inaf.it/cgi-bin/cmd

colours, which greatly limits the accuracy with which either can be estimated.

The radius and luminosity are estimated from the value of $T_{\rm eff}$ as determined from the *Gaia* photometry, including a bolometric correction obtained from synthetic spectra. The resulting estimates suffer from the naive use of $1/\varpi$ as a distance estimator and the assumption of zero extinction. Their uncertainties are probably underestimated.

6.3.5. Variability data

The variability data contained in *Gaia* DR2 is somewhat complex and consists of three data sets, as described in Sect. 3.4, that overlap to a large degree (for details refer to Holl et al. 2018). The mean G, $G_{\rm BP}$, and $G_{\rm RP}$ magnitudes and fluxes provided as part of the light curve statistics can differ from the values provided in the overall *Gaia* DR2 source table. In these cases the median or mean magnitudes and fluxes from the variability data set are to be preferred. There is a small number of stars with multiple entries in the SOS (Special Object Studies) tables and there are sources with a different type in the SOS and automated variability type estimation data sets. Classifications different from those of independent variable star surveys may occur (Holl et al. 2018; Arenou et al. 2018).

7. Using *Gaia* DR2 data: additional guidance

We briefly discuss a number of specific items that the users of *Gaia* DR2 should keep in mind. These concern issues inherent to the *Gaia* data (releases) and points to keep in mind when interpreting the results from analyses of *Gaia* DR2 data. More extensive examples of how to use the data responsibly are provided in the papers listed at the start of Sect. 4 and in Luri et al. (2018).

7.1. Time stamping in Gaia data releases

Gaia DR2 features photometric time series for sources varying in apparent magnitude and for solar system objects, as well as astrometric time series for the latter. Future releases will in addition contain time series for non-single star astrometry (such as binaries and stars with exoplanets), radial velocities, and the medium and low resolution spectra from the RVS and BP/RP instruments. As summarised in Lindegren et al. (2016) the primary coordinate system used for the *Gaia* (astrometric) data processing is the Barycentric Celestial Reference System (Soffel et al. 2003). The BCRS is a relativistic reference system that is physically adequate to describe both the motion of bodies in the solar system and the propagation of light from distant celestial sources. The time-like coordinate of the BCRS is the barycentric coordinate time (TCB). Consequently all the *Gaia* time series data are time-stamped using TCB. The numerical values in the *Gaia* DR2 tables are expressed JD−2455197.5(TCB) days, where by convention the origin for *Gaia* time-stamping is J2010.0(TCB) = JD 2455197.5(TCB).

7.2. Astrometric source model

All sources were treated as single stars in the astrometric solution for *Gaia* DR2 (Lindegren et al. 2018). This means that physical binaries and multiple systems as well as extended sources (galactic and extra-galactic, such as galaxies in the local universe), although present in *Gaia* DR2, received no special treatment. Moreover the sources that are not single stars are not marked as such. For binaries with orbital periods of the order of 2 yr the proper motions or parallaxes listed in *Gaia* DR2 may be quite far from the true values for the system. The auxiliary information in *Gaia* DR2 can be used to isolate candidate non-single stars or galaxies but this should be done with care and the results validated against known samples.

7.3. Solar system object astrometry

The epoch astrometry for SSOs is provided with uncertainties (on α, δ) and correlations. These correlations are strong, reflecting the large difference in precision between the along-scan and across-scan astrometric uncertainties which project into the uncertainties in (α, δ) in a correlated manner. The correlations should be taken into account for any application in order to recover the full accuracy of the astrometry in the along-scan direction. A known limitation of asteroid astrometry in *Gaia* DR2 is that the relativistic light deflection is computed as for the stars (i.e., the source is considered to be at infinite distance). A correction corresponding to the difference with respect to the finite distance must be applied whenever mas or sub-mas precision is aimed at.

7.4. Interpretation of photometric colours

The problem of the excess flux in the BP/RP photometry manifests itself primarily at the faint ($G > 19$) end of the survey, in crowded regions and around bright stars. In all these cases when constructing colour magnitude diagrams one should be careful in interpreting them.

For example, open cluster sequences in non-crowded fields may manifest a turn towards the blue at the lower end of the main sequence, which is a consequence of a stronger flux excess in BP than in RP for faint sources. At the faint end one should be aware that the effects of zodiacal light are clearly visible in the distribution of the flux excess factor (Evans et al. 2018).

Care should be taken in the use of colour magnitude diagrams in crowded regions such as globular cluster cores or the Milky Way bulge. Examples of colour-magnitude diagrams affected by the flux excess problem are given in Arenou et al. (2018). Finally, around bright sources there may be a dependence in source colour on the distance from the bright source which will lead to spurious features in a colour magnitude diagram.

When faint red sources are being analysed it may be better to use the $(G - G_{\rm RP})$ colour instead of $(G_{\rm BP} - G_{\rm RP})$ as discussed in Gaia Collaboration (2018a) for the case of brown dwarfs.

7.5. Mean magnitudes of variable stars

If a source is flagged as variable the recommendation is to use the mean value for its photometry from the tables with variability information, as the varying brightness of the source can be more carefully accounted for in the variability analysis.

7.6. Use the astrophysical parameters with care

Andrae et al. (2018) provide extensive guidance on the use of the astrophysical parameter estimates, including how to select samples with the most reliable $T_{\rm eff}$, radius, and luminosity estimates, and examples of how to use the estimates of A_G responsibly. We strongly recommend that these guidelines are followed and encourage independent investigations into the quality and limitations of the astrophysical parameter estimates.

7.7. Filtering to create clean samples

Although the bulk of the data in *Gaia* DR2 is of excellent quality, specific analyses of the data may require further filtering on data quality. One can find examples of how to do such filtering, using the information contained in *Gaia* DR2, throughout the papers accompanying the release. However, in many cases some experimentation by the user of the data will be needed to establish the best ad-hoc filtering for a given application. Such filtering does come at the cost of introducing additional truncation of the data which will further complicate the survey/sample selection function and may in fact severely bias the interpretation of the results. For example, Gaia Collaboration (2018d) show how a seemingly innocuous selection on radial velocity error can lead to strong kinematic biases when studying the Milky Way disk. Further examples of biases induced by sample truncation are given in Luri et al. (2018). Finally, one should keep in mind that filtering on the observed values or uncertainties of source parameters can increase the imprint on the resulting sample of, for example, scanning law patterns.

7.8. Negative parallaxes

Gaia DR2 represents the largest parallax catalogue ever produced and contains parallaxes of faint objects observed relatively few times and of extragalactic objects. For many of such objects the value of the parallax listed in the catalogue may be negative. As explained in Luri et al. (2018) the presence of negative parallaxes is a natural consequence of the way the *Gaia* observations are described in terms of a linearised astrometric source model, with the parameters of the model solved for through a least-squares process. Perhaps this is most easily appreciated by considering the 0.5 million QSOs appearing in *Gaia* DR2 for which parallax solutions have been made. Given that the true parallax for these sources is close to zero it is to be expected that for half of them the observed parallax (as solved for from the observations) is negative (where in the case of *Gaia* DR2 the fraction of negative parallaxes for QSOs is higher because of the negative parallax zeropoint).

Hence negative parallaxes represent perfectly valid measurements and can be included in analyses of the *Gaia* DR2 data. Examples of how one can do this are given in Luri et al. (2018).

7.9. Known spurious results

There are a number of results listed in *Gaia* DR2 which are obviously wrong and which may surprise the user of the data. We point out two specific cases here.

For a small number of sources the parallaxes listed in *Gaia* DR2 have very large positive or negative values (with for example 59 sources having parallaxes larger than that of Proxima Centauri), where the negative values can be very far from zero when expressed in terms of the formal uncertainty on the parallax. These parallax values are spurious and caused by a close alignment (of order 0.2–0.3 arcsec) of sources, that are only occasionally resolved in the *Gaia* observations, depending on the scan direction. These cases show up typically in dense regions covered by only a few transits or an unfortunate distribution of scan directions and parallax factors. This is consistent with most of these sources being faint and concentrated in dense areas along the Galactic plane and toward the Galactic bulge (see Fig. C4 in Lindegren et al. 2018). Most likely the proper motions of these sources are also erroneous. This is consistent with the presence of a number of high-proper motion stars at $G > 19$ (104 243 at $\mu > 100$ mas yr^{-1}, 12 431 at $\mu > 200$ mas yr^{-1}, and 4459 at $\mu > 300$ mas yr^{-1}) which show a marked concentration toward the galactic bulge and galactic plane regions. These sources overlap to a large degree with the sources with spurious parallax values and their proper motions are thus likely to be unreliable. More details on this problem and guidance on how to clean samples from spurious parallaxes can be found in Lindegren et al. (2018, in particular their Appendix C).

Among the bright and well known (i.e. named) variable stars there are a number of cases where the mean G-band magnitude listed in *Gaia* DR2 is clearly wrong. One prominent case is the star RR Lyrae itself for which the mean magnitude is listed as $G = 17$. The wrong value is caused by the fact that the treatment of outliers, as implemented in the photometric processing for *Gaia* DR2, is not efficient in the case of variable sources that have an intrinsically large spread in the individual photometric observations. As a consequence of the wrong magnitude estimate, the parallax of RR Lyrae was determined to be −2.6 mas.

We stress here that the above problems concern only a very small number of cases which do not indicate overall problems with the quality of *Gaia* DR2.

7.10. Take into account uncertainties and correlations

The astrometric uncertainties are provided in the form of the full covariance matrix for the five astrometric parameters. The correlations between the uncertainties can be significant and they should always be accounted for to correctly calculate the standard uncertainties on linear combinations of (subsets of) the astrometric parameters and to correctly assess, for example, how far away a given set of astrometric parameters is from a model prediction. The mathematics involved in accounting for correlated uncertainties is summarised in Luri et al. (2018) and described more extensively in the *Gaia* DR2 online documentation.

In this context we point out that the longest principal axis of a scaled version of the covariance matrix is provided as the parameter `astrometric_sigma_5dmax` for both the 5-parameter and 2-parameter solutions. This parameter is equivalent to the semi-major axis of the position error ellipse and can be useful in filtering out sources for which one of the astrometric parameters, or a linear combination of several parameters, is particularly ill-determined. We refer to the online *Gaia* DR2 documentation for more details.

7.11. Dealing with underestimated uncertainties and/or systematic errors

As pointed out above the uncertainties quoted in *Gaia* DR2 on the various source parameters can be underestimated and there are also systematic errors with varying dependencies on source brightness, colour, and position on the sky, which moreover may be spatially correlated. We can provide no general recipe for taking these effects into account in scientific analyses of the *Gaia* DR2 data, but give a few recommendations here.

We strongly advise against attempts to "correct" the data themselves as a means to get rid of underestimated uncertainties or systematic errors. This would require a level of understanding and characterisation of these effects that would have allowed their removal during the data processing in the first place. We recommend (for studies where it matters) to include the presence of systematic effects in the uncertainties as part of the data analysis, for example in a forward modelling approach. The level of systematic errors (e.g. the size of the parallax zeropoint) or

the factors by which uncertainties are under- or overestimated then become part of the model parameters to estimate. Examples of such analyses of *Gaia* DR1 parallax data can be found in Casertano et al. (2017) and Sesar et al. (2017), where the latter include both a parallax zeropoint and a scaling factor for the quoted uncertainties as part of their probabilistic model that fits a period luminosity relation to data for RR Lyrae stars. The spatial correlation parameters for the uncertainties and systematic errors can be included in a similar way as part of the modelling. Further guidance on the use of the astrometric data (in particular the parallaxes) from *Gaia* DR2 can be found in Luri et al. (2018).

8. *Gaia* DR2 access facilities

The main entry point to *Gaia* DR2 remains the ESA *Gaia* archive[4]. Access is also possible through a number of partner and affiliate data centres in Europe, the United States, Japan, Australia, and South Africa. These data centres provide their own access facilities, but do not necessarily host all data contained in the ESA *Gaia* archive. The services offered at the ESA *Gaia* archive remain as described in Gaia Collaboration (2016a) and we list here a few enhancements and changes:

- The access to the light curves for variable stars is now in the form of a URL that links from the main `gaia_source` table to the specific files that contain the light curves for the source in VOTable format[5].
- The astrometric and photometric time series for the SSOs are all collated into one large table containing multiple entries for each SSO. Note that the source identifiers for SSOs are negative numbers. To enable queries of SSOs based on orbital elements or absolute magnitude, an auxiliary table containing such data, plus ancillary quantities, is provided. In addition a table with the residuals of each *Gaia* observation with respect to an orbital fit is provided as a reference.
- The archive visualisation service (Moitinho et al. 2017) has been much expanded to allow for efficient preliminary exploration of the data in the entire *Gaia* DR2 catalogue. The service offers several pre-computed diagrams which can be explored through linked views and allows one to interactively define a query for a given data set. This serves in particular to narrow down queries for data to the exact samples one is interested in and thus save time and storage space for the actual query. Full details can be found in Moitinho et al. (2018).
- We provide pre-computed cross-matches between *Gaia* DR2 and a number of other large surveys. We recommend using these cross-matches as they have been carefully validated and their use facilitates reproducing analyses of *Gaia* DR2 data combined with other survey data. The details are provided in Marrese et al. (2018). The pre-computed cross-matches are provided for the following surveys: HIPPARCOS (new reduction, van Leeuwen 2007); *Tycho*-2 (Høg et al. 2000); 2MASS (Skrutskie et al. 2006); SDSS DR9 (Ahn et al. 2012); APASS DR9 (Henden et al. 2016, 2015); UCAC4 (Zacharias et al. 2013); Pan-STARRS1 (Chambers et al. 2016); AllWise (Wright et al. 2010); GSC2.3 (Lasker et al. 2008); PPMXL (Roeser et al. 2010); URAT1 (Zacharias et al. 2015); and RAVE DR5 (Kunder et al. 2017).

[4] It can be accessed at `http://archives.esac.esa.int/gaia`
[5] `http://www.ivoa.net/documents/latest/VOT.html`

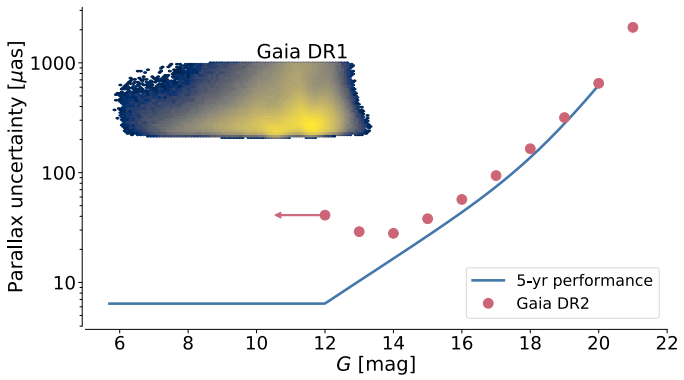

Fig. 7. Parallax uncertainties in *Gaia* DR2 (dots) as a function of G compared to the uncertainties quoted for *Gaia* DR1 (colour scale) and the expected end-of-mission parallax performance (solid line), as predicted after the commissioning of *Gaia*. Note how the performance for *Gaia* DR2 is still limited by calibration uncertainties for sources brighter than $G \sim 14$.

Finally we mention the creation of a *Gaia* Community forum[6] which is intended to facilitate discussion on the use of *Gaia* data. The principle is to let the users of the data discuss amongst themselves on this forum but the discussions will be monitored by members from the *Gaia* Data Processing and Analysis Consortium who may respond with comments and expert advice when necessary.

9. Conclusions

With the first *Gaia* data release in 2016 the astronomical community got an early taste of the potential of the *Gaia* mission results, in particular through the 2 million parallaxes and proper motions made available as part of the *Tycho-Gaia* Astrometric Solution. The science done with *Gaia* DR1 spans a wide range of topics and often features the powerful combination of *Gaia* and other surveys. *Gaia* DR1 was also quickly established as a reference for the astrometric and photometric calibration of other surveys, resulting among others in the rejuvenation of existing proper motion catalogues. For a brief review of the impact of *Gaia* DR1 we refer to Brown (2017).

With the release of *Gaia* DR2 the promise of the availability of fundamental astrophysical information for (over) a billion sources spread over a substantial fraction of the volume of the Milky Way starts to be fulfilled. The addition of the largest radial velocity survey to date, coupled with astrophysical information for 161 million sources and variability information for half a million sources will make *Gaia* DR2 a resource to be mined for stellar physics and galactic as well as extra-galactic astronomy for many years to come. Moreover, *Gaia* DR2 provides a first glimpse of the immense power of *Gaia* for solar system studies.

Nevertheless *Gaia* DR2 still represents an early data release based on only a limited amount (less than two years) of input data, partly inadequate calibrations, and an incomplete understanding of the behaviour of the spacecraft, payload, and instruments. These shortcomings manifest themselves as systematic errors which although much reduced in size from *Gaia* DR1 to *Gaia* DR2 will remain a limiting factor in scientific uses of the data, in particular at the bright end of the survey and, for example, for distant samples. This is illustrated in Fig. 7 which shows the parallax uncertainties as a function of G for *Gaia* DR2 (dots),

[6] `https://www.cosmos.esa.int/web/gaia/forum`

Gaia DR1 (colour scale map), and the end of mission (solid line, as predicted after *Gaia* commissioning, Gaia Collaboration 2016b). The bright end ($G \lesssim 14$) performance for *Gaia* DR2 is still limited by calibration errors, while at the faint end the nominal end of mission performance is already being reached (this is probably due to a conservative assessment of the effect at the faint end of the excess stray light). The task for the *Gaia* data processing for the next data release will thus be to substantially reduce the systematics such that a real advantage can be gained, in particular at the bright end, from the increase in precision due to the longer time span of the input data. The main challenges will be the following. The PSF modelling used in the image location determination must be upgraded, such that for example astrometric colour terms are already accounted for at an early stage. This also requires the closing of the iterative loop shown in Fig. 10 in Gaia Collaboration (2016a). The modelling of the sky background (both astronomical and as caused by the excess stray light on board *Gaia*) has to be refined to further improve the image location process and to get rid of the flux excess in the BP/RP photometry. The latter will also benefit from an improvement in the treatment of crowded fields, specifically a better treatment of the effects of overlapping images in all of *Gaia*'s instruments and in particular for the BP/RP/RVS instruments where the measurement of spectra necessitates much larger images in the focal plane. Finally the origins of the systematic effects in the astrometry will be further investigated with much effort to be dedicated to the continued development of the possibility to calibrate the systematic effects from the observations.

The next *Gaia* data release will also feature new data products of which the BP/RP and RVS spectra and the non-single star astrometric and radial velocity solutions represent qualitative changes in the character with respect to *Gaia* DR2. Further enhancements include: epoch astrometry for non-single stars, an expanded radial velocity survey (to $G_{\rm RVS} \sim 14$) including the analysis of spectroscopic binaries, astrophysical parameter estimates based on BP/RP/RVS spectra, a further order of magnitude increase in the availability of variability information, the first results from eclipsing binary star processing, analyses of extended objects (galaxies, QSO hosts), and an expanded list of some hundred thousand solar system objects for which multi-colour photometry will also be provided. The latter opens up for investigation the powerful combination of precise orbits for SSOs combined with a homogeneous multi-colour photometric survey of these bodies.

Hence there is much more to come from *Gaia* but for now we invite the reader to start exploring the magnificent survey that is *Gaia* DR2.

Acknowledgements. This work presents results from the European Space Agency (ESA) space mission *Gaia*. *Gaia* data are being processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the *Gaia* MultiLateral Agreement (MLA). The *Gaia* mission website is https://www.cosmos.esa.int/gaia. The *Gaia* Archive website is http://gea.esac.esa.int/archive/. The *Gaia* mission and data processing have financially been supported by, in alphabetical order by country: the Algerian Centre de Recherche en Astronomie, Astrophysique et Géophysique of Bouzareah Observatory; the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF) Hertha Firnberg Programme through grants T359, P20046, and P23737; the BELgian federal Science Policy Office (BELSPO) through various PROgramme de Développement d'EXpériences scientifiques (PRODEX) grants and the Polish Academy of Sciences - Fonds Wetenschappelijk Onderzoek through grant VS.091.16N; the Brazil-France exchange programmes Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Coordenação de Aperfeicoamento de Pessoal de Nível Superior (CAPES) - Comité Français d'Evaluation de la Coopération Universitaire et Scientifique avec le Brésil (COFECUB); the Chilean Dirección de Gestión de la Investigación (DGI) at the University of Antofagasta and the Comité Mixto ESO-Chile; the National Science Foundation of China (NSFC) through grants 11573054 and 11703065; the Czech-Republic Ministry of Education, Youth, and Sports through grant LG 15010, the Czech Space Office through ESA PECS contract 98058, and Charles University Prague through grant PRIMUS/SCI/17; the Danish Ministry of Science; the Estonian Ministry of Education and Research through grant IUT40-1; the European Commission's Sixth Framework Programme through the European Leadership in Space Astrometry (ELSA) Marie Curie Research Training Network (MRTN-CT-2006-033481), through Marie Curie project PIOF-GA-2009-255267 (Space AsteroSeismology & RR Lyrae stars, SAS-RRL), and through a Marie Curie Transfer-of-Knowledge (ToK) fellowship (MTKD-CT-2004-014188); the European Commission's Seventh Framework Programme through grant FP7-606740 (FP7-SPACE-2013-1) for the *Gaia* European Network for Improved data User Services (GENIUS) and through grant 264895 for the *Gaia* Research for European Astronomy Training (GREAT-ITN) network; the European Research Council (ERC) through grants 320360 and 647208 and through the European Union's Horizon 2020 research and innovation programme through grants 670519 (Mixing and Angular Momentum tranSport of massIvE stars – MAMSIE) and 687378 (Small Bodies: Near and Far); the European Science Foundation (ESF), in the framework of the *Gaia* Research for European Astronomy Training Research Network Programme (GREAT-ESF); the European Space Agency (ESA) in the framework of the *Gaia* project, through the Plan for European Cooperating States (PECS) programme through grants for Slovenia, through contracts C98090 and 4000106398/12/NL/KML for Hungary, and through contract 4000115263/15/NL/IB for Germany; the European Union (EU) through a European Regional Development Fund (ERDF) for Galicia, Spain; the Academy of Finland and the Magnus Ehrnrooth Foundation; the French Centre National de la Recherche Scientifique (CNRS) through action "Défi MASTODONS", the Centre National d'Etudes Spatiales (CNES), the L'Agence Nationale de la Recherche (ANR) "Investissements d'avenir" Initiatives D'EXcellence (IDEX) programme Paris Sciences et Lettres (PSL$*$) through grant ANR-10-IDEX-0001-02, the ANR "Défi de tous les savoirs" (DS10) programme through grant ANR-15-CE31-0007 for project "Modelling the Milky Way in the Gaia era" (MOD4Gaia), the Région Aquitaine, the Université de Bordeaux, and the Utinam Institute of the Université de Franche-Comté, supported by the Région de Franche-Comté and the Institut des Sciences de l'Univers (INSU); the German Aerospace Agency (Deutsches Zentrum für Luft- und Raumfahrt e.V., DLR) through grants 50QG0501, 50QG0601, 50QG0602, 50QG0701, 50QG0901, 50QG1001, 50QG1101, 50QG1401, 50QG1402, 50QG1403, and 50QG1404 and the Centre for Information Services and High Performance Computing (ZIH) at the Technische Universität (TU) Dresden for generous allocations of computer time; the Hungarian Academy of Sciences through the Lendület Programme LP2014-17 and the János Bolyai Research Scholarship (L. Molnár and E. Plachy) and the Hungarian National Research, Development, and Innovation Office through grants NKFIH K-115709, PD-116175, and PD-121203; the Science Foundation Ireland (SFI) through a Royal Society - SFI University Research Fellowship (M. Fraser); the Israel Science Foundation (ISF) through grant 848/16; the Agenzia Spaziale Italiana (ASI) through contracts I/037/08/0, I/058/10/0, 2014-025-R.0, and 2014-025-R.1.2015 to the Italian Istituto Nazionale di Astrofisica (INAF), contract 2014-049-R.0/1/2 to INAF dedicated to the Space Science Data Centre (SSDC, formerly known as the ASI Science Data Centre, ASDC), and contracts I/008/10/0, 2013/030/I.0, 2013-030-I.0.1-2015, and 2016-17-I.0 to the Aerospace Logistics Technology Engineering Company (ALTEC S.p.A.), and INAF; the Netherlands Organisation for Scientific Research (NWO) through grant NWO-M-614.061.414 and through a VICI grant (A. Helmi) and the Netherlands Research School for Astronomy (NOVA); the Polish National Science Centre through HARMONIA grant 2015/18/M/ST9/00544 and ETIUDA grants 2016/20/S/ST9/00162 and 2016/20/T/ST9/00170; the Portugese Fundação para a Ciência e a Tecnologia (FCT) through grant SFRH/BPD/74697/2010; the Strategic Programmes UID/FIS/00099/2013 for CENTRA and UID/EEA/00066/2013 for UNINOVA; the Slovenian Research Agency through grant P1-0188; the Spanish Ministry of Economy (MINECO/FEDER, UE) through grants ESP2014-55996-C2-1-R, ESP2014-55996-C2-2-R, ESP2016-80079-C2-1-R, and ESP2016-80079-C2-2-R, the Spanish Ministerio de Economía, Industria y Competitividad through grant AyA2014-55216, the Spanish Ministerio de Educación, Cultura y Deporte (MECD) through grant FPU16/03827, the Institute of Cosmos Sciences University of Barcelona (ICCUB, Unidad de Excelencia "María de Maeztu") through grant MDM-2014-0369, the Xunta de Galicia and the Centros Singulares de Investigación de Galicia for the period 2016-2019 through the Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC), the Red Española de Supercomputación (RES) computer resources at MareNostrum, and the Barcelona Supercomputing Centre - Centro Nacional de Supercomputación (BSC-CNS) through activities AECT-2016-1-0006, AECT-2016-2-0013, AECT-2016-3-0011, and AECT-2017-1-0020; the Swedish National Space Board (SNSB/Rymdstyrelsen); the Swiss State Secretariat for Education, Research, and Innovation through the ESA PRODEX programme, the

Mesures d'Accompagnement, the Swiss Activités Nationales Complémentaires, and the Swiss National Science Foundation; the United Kingdom Rutherford Appleton Laboratory, the United Kingdom Science and Technology Facilities Council (STFC) through grant ST/L006553/1, the United Kingdom Space Agency (UKSA) through grant ST/N000641/1 and ST/N001117/1, as well as a Particle Physics and Astronomy Research Council Grant PP/C503703/1. The GBOT programme (Gaia Collaboration 2016b; Altmann et al. 2014) uses observations collected at (i) the European Organisation for Astronomical Research in the Southern Hemisphere (ESO) with the VLT Survey Telescope (VST), under ESO programmes 092.B-0165, 093.B-0236, 094.B-0181, 095.B-0046, 096.B-0162, 097.B-0304, 098.B-0034, 099.B-0030, 0100.B-0131, and 0101.B-0156, and (ii) the Liverpool Telescope, which is operated on the island of La Palma by Liverpool John Moores University in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias with financial support from the United Kingdom Science and Technology Facilities Council, and (iii) telescopes of the Las Cumbres Observatory Global Telescope Network. In this work we made use of the Set of Identifications, Measurements, and Bibliography for Astronomical Data (SIMBAD; Wenger et al. 2000), the "Aladin sky atlas" (Bonnarel et al. 2000; Boch & Fernique 2014), and the VizieR catalogue access tool (Ochsenbein et al. 2000), all operated at the Centre de Données astronomiques de Strasbourg (CDS). We additionally made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2018), IPython (Pérez & Granger 2007), Matplotlib (Hunter 2007), and TOPCAT (Taylor 2005, http://www.starlink.ac.uk/topcat/). We thank the anonymous referee for suggestions that greatly helped improve the readability and clarity of this paper.

References

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21
Altmann, M., Bouquillon, S., Taris, F., et al. 2014, in Observatory Operations: Strategies, Processes, and Systems V, Proc. SPIE, 9149, 91490P
Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, A&A, 616, A8 (*Gaia* 2 SI)
Arenou, F., Luri, X., Babusiaux, C., et al. 2017, A&A, 599, A50
Arenou, F., Luri, X., Babusiaux, C., et al. 2018, A&A, 616, A17 (*Gaia* 2 SI)
Astropy Collaboration (Price-Whelan, et al.) 2018, ArXiv e-prints [arXiv:1801.02634]
Boch, T., & Fernique, P. 2014, in Astronomical Data Analysis Software and Systems XXIII, eds. N. Manset & P. Forshay, ASP Conf. Ser., 485, 277
Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, A&AS, 143, 33
Brown, A. G. A. 2017, Proc. IAU Symp. 330, 13
Casertano, S., Riess, A. G., Bucciarelli, B., & Lattanzi, M. G. 2017, A&A, 599, A67
Castelli, F., & Kurucz, R. L. 2003, Proc. IAU Symp. 210, poster A20 [arXiv:astro-ph/0405087]
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:1612.05560]
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Cropper, M., Katz, D., Sartoretti, P., et al. 2018, A&A, 616, A5 (*Gaia* 2 SI)
Evans, D. W., Riello, M., De Angeli, F., et al. 2017, A&A, 600, A51
Evans, D. W.,, Riello, M., De Angeli, F., et al. 2018, A&A, 616, A4 (*Gaia* 2 SI)
Fabricius, C., Bastian, U., Portell, J., et al. 2016, A&A, 595, A3
Gaia Collaboration (Brown, A. G. A., et al.) 2016a, A&A, 595, A2
Gaia Collaboration (Prusti, T., et al.) 2016b, A&A, 595, A1
Gaia Collaboration (Babusiaux, C., et al.) 2018a, A&A, 616, A10 (*Gaia* 2 SI)
Gaia Collaboration (Eyer, L., et al.) 2018b, A&A, submitted (*Gaia* 2 SI)
Gaia Collaboration (Helmi, A., et al.) 2018c, A&A, 616, A12 (*Gaia* 2 SI)
Gaia Collaboration (Katz, D., et al.) 2018d, A&A, 616, A11 (*Gaia* 2 SI)
Gaia Collaboration (Mignard, F., et al.) 2018e, A&A, 616, A14 (*Gaia* 2 SI)
Gaia Collaboration (Spoto, F., et al.) 2018f, A&A, 616, A13 (*Gaia* 2 SI)
Geurts, P., Ernst, D., & Wehenkel, L. 2006, Mach. Learn., 63, 3
Hambly, N., Cropper, M., Boudreault, S., et al. 2018, A&A, 616, A15 (*Gaia* 2 SI)
Hawkins, K., Leistedt, B., Bovy, J., & Hogg, D. W. 2017, MNRAS, 471, 722
Henden, A. A., Levine, S., Terrell, D., & Welch, D. L. 2015, AAS Meeting Abstracts, 225, 336.16
Henden, A. A., Templeton, M., Terrell, D., et al. 2016, VizieR Online Data Catalogue, II/336
Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27
Holl, B., Audard, M., Nienartowicz, K., et al. 2018, A&A, in press, DOI: 10.1051/0004-6361/201832892 (*Gaia* 2 SI)
Hunter, J. D. 2007, Comput. Sci. Eng.,, 9, 90
Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
Katz, D., Sartoretti, P., Cropper, M., et al. 2018, A&A, submitted (*Gaia* 2 SI)
Kunder, A., Kordopatis, G., Steinmetz, M., et al. 2017, AJ, 153, 75
Lasker, B. M., Lattanzi, M. G., McLean, B. J., et al. 2008, AJ, 136, 735
Lindegren, L., Lammers, U., Bastian, U., et al. 2016, A&A, 595, A4
Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2 (*Gaia* 2 SI)
Luri, X., Brown, A., Sarro, L., et al. 2018, A&A, 616, A9
Maíz Apellániz, J. 2017, A&A, 608, L8
Makarov, V. V., Fabricius, C., & Frouard, J. 2017, ApJ, 840, L1
Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
Marrese, P., Marinoni, S., Fabrizio, M., & Altavilla, G. 2018, A&A, submitted (*Gaia* 2 SI)
Michalik, D., Lindegren, L., & Hobbs, D. 2015a, A&A, 574, A115
Michalik, D., Lindegren, L., Hobbs, D., & Butkevich, A. G. 2015b, A&A, 583, A68
Moitinho, A., Krone-Martins, A., Savietto, H., et al. 2017, A&A, 605, A52
Moitinho, A. et al. 2018, A&A, submitted (*Gaia* 2 SI)
Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23
Pérez, F., & Granger, B. E. 2007, Comput. Sci. Eng., 9, 21
Riello, M., De Angeli, F., Evans, D., et al. 2018, A&A, 616, A3 (*Gaia* 2 SI)
Roeser, S., Demleitner, M., & Schilbach, E. 2010, AJ, 139, 2440
Ruiz-Dern, L., Babusiaux, C., Arenou, F., Turon, C., & Lallement, R. 2018, A&A, 609, A116
Sartoretti, P., Katz, D., Cropper, M., et al. 2018, A&A, 616, A6 (*Gaia* 2 SI)
Sesar, B., Fouesneau, M., Price-Whelan, A. M., et al. 2017, ApJ, 838, 107
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Smart, R. L., & Nicastro, L. 2014, A&A, 570, A87
Soffel, M., Klioner, S. A., Petit, G., et al. 2003, AJ, 126, 2687
Soubiran, C., Jasniewicz, G., Chemin, L., et al. 2018, A&A, 616, A7 (*Gaia* 2 SI)
Taylor, M. B. 2005, in Astronomical Data Analysis Software and Systems XIV, eds. P. Shopbell, M. Britton, & R. Ebert, ASP Conf. Ser., 347, 29
van Leeuwen, F. 2007, Astrophys. Space Sci. Lib., 350
Weiler, M., Jordi, C., Fabricius, C., & Carrasco, J. M. 2018, A&A, 615, A24
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Zacharias, N., Finch, C. T., Girard, T. M., et al. 2013, AJ, 145, 44
Zacharias, N., Finch, C., Subasavage, J., et al. 2015, AJ, 150, 101

[1] Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
[2] INAF - Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122 Padova, Italy
[3] Science Support Office, Directorate of Science, European Space Research and Technology Centre (ESA/ESTEC), Keplerlaan 1, 2201AZ, Noordwijk, The Netherlands
[4] GEPI, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France
[5] Univ. Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France
[6] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[7] Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstr. 12-14, 69120 Heidelberg, Germany
[8] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[9] Department of Astronomy, University of Geneva, Chemin des Maillettes 51, 1290 Versoix, Switzerland
[10] Mission Operations Division, Operations Department, Directorate of Science, European Space Research and Technology Centre (ESA/ESTEC), Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
[11] Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
[12] Lohrmann Observatory, Technische Universität Dresden, Mommsenstraße 13, 01062 Dresden, Germany
[13] European Space Astronomy Centre (ESA/ESAC), Camino bajo del Castillo s/n, Urbanizaciòn Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[14] Lund Observatory, Department of Astronomy and Theoretical Physics, Lund University, Box 43, 22100 Lund, Sweden
[15] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
[16] CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France