The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/162635

Please be advised that this information was generated on 2019-03-03 and may be subject to change.
Gaia Data Release 1

Summary of the astrometric, photometric, and survey properties


(Affiliations can be found after the references)

Received 10 August 2016 / Accepted 31 August 2016

ABSTRACT

Context. At about 1000 days after the launch of Gaia we present the first Gaia data release, Gaia DR1, consisting of astrometry and photometry for over 1 billion sources brighter than magnitude 20.7.

Aims. A summary of Gaia DR1 is presented along with illustrations of the scientific quality of the data, followed by a discussion of the limitations due to the preliminary nature of this release.

Methods. The raw data collected by Gaia during the first 14 months of the mission have been processed by the Gaia Data Processing and Analysis Consortium (DPAC) and turned into an astrometric and photometric catalogue.

Results. Gaia DR1 consists of three components: a primary astrometric data set which contains the positions, parallaxes, and proper motions for about 2 million of the brightest stars in common with the HIPPARCOS and Tycho-2 catalogues – a realisation of the Tycho-Gaia Astrometric Solution (TGAS) – and a secondary astrometric data set containing the positions for an additional 1.1 billion sources. The second component is the photometric data set, containing mean G-band magnitudes for all sources. The G-band light curves and the characteristics of ~3000 Cepheid and RR Lyrae stars, observed at high cadence around the south ecliptic pole, form the third component. For the primary astrometric data set the typical uncertainty is about 0.3 mas for the positions and parallaxes, and about 1 mas yr^-1 for the proper motions. A systematic component of ~0.3 mas should be added to the parallax uncertainties. For the subset of ~94 000 HIPPARCOS stars in the primary data set, the proper motions are much more precise at about 0.06 mas yr^-1. For the secondary astrometric data set, the typical uncertainty of the positions is ~10 mas. The median uncertainties on the mean G-band magnitudes range from the mmag level to ~0.03 mag over the magnitude range 5 to 20.7.

Conclusions. Gaia DR1 is an important milestone ahead of the next Gaia data release, which will feature five-parameter astrometry for all sources. Extensive validation shows that Gaia DR1 represents a major advance in the mapping of the heavens and the availability of basic stellar data that underpin observational astrophysics. Nevertheless, the very preliminary nature of this first Gaia data release does lead to a number of important limitations to the data quality which should be carefully considered before drawing conclusions from the data.

Key words catalogues – astrometry – parallaxes – proper motions – surveys
1. Introduction

The Gaia satellite was launched at the end of 2013 to collect data that will allow the determination of highly accurate positions, parallaxes, and proper motions for >1 billion sources brighter than magnitude 20.7 in the white-light photometric band $G$ of Gaia (thus going deeper than the originally planned limit of $G = 20$). The astrometry is complemented by multi-colour photometry, measured for all sources observed by Gaia, and radial velocities which are collected for stars brighter than $G \approx 17$. The scientific goals of the mission are summarised in Gaia Collaboration (2016b), while a more extensive scientific motivation for the mission is presented in Perryman et al. (2001).

The spacecraft, its scientific instruments, and the observing strategy have been designed to meet the performance requirement of 24 µas accuracy on the parallax of a 15th magnitude solar-type star at the end of the nominal 5 yr mission lifetime. The entity entrusted with the data processing for the Gaia mission, the Gaia Data Processing and Analysis Consortium (DPAC, described in Gaia Collaboration 2016b), is expected to deliver the final data products (at their ultimately achievable accuracy) only at the end of post-operational phase of the mission, currently foreseen for 2022–2023. It was therefore agreed at the time of the creation of DPAC that the astronomical community should have access to the Gaia data at an earlier stage through intermediate data releases. It was understood that these intermediate releases are based on preliminary calibrations and only on a subset of the measurements available at the end of the mission, and therefore will not be representative of the end-of-mission Gaia performance.

In this paper we present the first such intermediate Gaia data release (Gaia Data Release 1, Gaia DR1), which is based on the data collected during the first 14 months of the nominal mission lifetime (60 months). In Sect. 2 we provide a short summary of the Gaia instruments and the way the data are collected. We summarise the astrometric, photometric and variable star contents of Gaia DR1 in Sect. 3. A summary of the validation of the results is provided in Sect. 4 and a few illustrations of the contents of Gaia DR1 are provided in Sect. 5. The known limitations of this first release are presented in Sect. 6. In Sect. 7 we provide pointers to the Gaia DR1 data access facilities and documentation available to the astronomical community. We conclude in Sect. 8. Although Gaia DR1 is the first major catalogue release with results from the Gaia mission, Gaia data has already been made publicly available as “Science Alerts” on transient sources, which for example led to the discovery of only the third known eclipsing AM CVn-system (Campbell et al. 2015).

We stress at the outset that Gaia DR1 represents a preliminary release of Gaia results with many shortcomings. We therefore strongly encourage a detailed reading of Sect. 6 and the documentation associated with the release as well as carefully taking into account the listed limitations when drawing conclusions based on the data contained in Gaia DR1.

2. Gaia instruments and measurements

We provide a brief overview of the Gaia instruments and the way measurements are collected in order to introduce some of the technical terms used in the rest of the paper. A full description of the Gaia spacecraft, instruments, and measurement principles can be found in Gaia Collaboration (2016b).

Gaia continuously scans the sky with two telescopes pointing in directions separated by the basic angle of 106.5°. The images produced by the telescopes are projected onto the same focal plane composed of 106 CCDs which function as the detectors of the various instruments in the Gaia payload. The scanning is achieved through the continuous revolution of Gaia about its spin axis with a period of 6 h. The spin axis direction precesses around the direction to the Sun (as seen from Gaia), which allows complete coverage of the sky. Statistics of the sky coverage achieved for Gaia DR1 are presented in Lindegren et al. (2016) and van Leeuwen et al. (2016), while the properties of the Gaia scanning law with respect to variable star studies are described in Eyner et al. (2016).

The spinning motion of the spacecraft results in the source images moving across the focal plane. This necessitates the operation of the Gaia CCDs in time-delayed integration (TDI) mode so as to allow the accumulation of charge as the images move across the CCDs. The CCDs are not fully read out, only the pixels in a “window” around each source are read out and stored for transmission to the ground. These windows come in various sizes and sampling schemes.

The astrometric instrument takes up most of the focal plane and collects source images in the Gaia white-light pass band $G$ (covering the range 330–1050 nm, Carrasco et al. 2016; Jordi et al. 2010). The fundamental inputs to the astrometric data processing consist of the precise times when the image centroids pass a fiducial line on the CCD (Lindegren et al. 2012). The image centroid and the flux contained in the image are determined as part of the pre-processing (Fabricius et al. 2016). The sensitivity of the astrometric instrument is such that sources brighter than about $G \sim 12$ will lead to saturated images. This effect is mitigated through the use of TDI gates, which are special structures on the CCDs that can be activated to inhibit charge transfer and hence to effectively reduce the integration time for bright sources.

The photometric instrument is realised through two prisms dispersing the light entering the field of view of two dedicated sets of CCDs. The Blue Photometer (BP) operates over the wavelength range 330–680 nm, while the Red Photometer (RP) covers the wavelength range 640–1050 nm (Carrasco et al. 2016; Jordi et al. 2010). The data collected by the photometric instrument consists of low resolution spectrophotometric measurements of the source spectral energy distributions. This colour information is intended for use in the astrometric processing (to correct for chromatic effects) and to provide the astrophysical characterisation of all sources observed by Gaia. The G-band photometry is derived from the fluxes measured in the astrometric instrument. Results from the photometric instrument are not presented as part of Gaia DR1. The photometry in this first release only concerns the fluxes measured in the G band.

The spectroscopic instrument, also called the radial-velocity spectrometer (RVS) collects medium resolution ($R \sim 11,700$) spectra over the wavelength range 845–872 nm, centred on the Calcium triplet region (Cropper & Katz 2011). The spectra are collected for all sources to $G \approx 17$ (16th magnitude in the RVS filter band) and serve primarily to determine the radial velocity of the sources, although at the bright end ($G < 12.5$, Recio-Blanco et al. 2016) astrophysical information can be derived directly from the spectra. Results from this instrument are not contained in Gaia DR1.

Observations of sources by Gaia can be referred to in several ways. “Focal plane transits” refer to a crossing of the entire focal plane by a given source, which corresponds to a “visit” by Gaia of a specific coordinate on the sky. “CCD transits” refer to the crossing by a source of a particular CCD in the focal plane. Thus the focal plane transit of the astrometric field typically consists of 10 transits across individual CCDs, while a photometric
instrument transit (BP or RP) consists of only one CCD transit, and a transit across the RVS instrument consists of three CCD transits (see Gaia Collaboration 2016b; Crowley et al. 2016b, for more details on the focal plane layout and functionalities, and the in-flight performance of the Gaia CCDs). This distinction is important when it comes to the difference between the number of measurements (CCD transits) collected for a source and the number of times it was observed (focal plane transits) by Gaia.

In the rest of the paper we will refer to an “observation” or a “focal plane transit” to indicate that a source was observed by Gaia and we refer to “CCD transit” whenever individual CCD measurements are discussed.

Events on board Gaia are labelled by the so-called on-board mission time line (OBMT), which is a time scale defined by the on-board clock. This time scale is eventually transformed into the physical barycentric coordinate time (TCB) (Gaia Collaboration 2016b; Lindegren et al. 2016). By convention OBMT is expressed in units of 0 s (21 600 s) spacecraft revolutions since launch and this unit is often used in figures of some quantity versus time, including in the papers accompanying Gaia DR1 and in the data release documentation (see Sect. 7). For the practical interpretation of time lines expressed in OBMT the following approximate relation between the OBMT (in revolutions) and TCB at Gaia (in Julian years) can be used:

\[ \text{TCB} \approx J2015.0 + \left( \text{OBMT} - 1717.6256 \text{ rev} \right) / \left(1461 \text{ rev} \right). \]  

This relation is precise to ±2 s and is valid only for the time span corresponding to Gaia DR1. The time interval covered by the observations used for Gaia DR1 starts at OBMT 1078.3795 rev = J2014.5624599 TCB (approximately 2014 July 25, 10:30:00 UTC), and ends at OBMT 2751.3518 rev = J2015.7075471 TCB (approximately 2015 September 16, 16:20:00 UTC), thus spanning 418 days. This time interval contains a significant number of gaps which are caused by: events or operations on board Gaia that prevent the collection of data or make the raw data unusable for a while (such as the decontamination of the payload, problems in the pre-processing leading to effective gaps in the available raw Gaia data (which has to be reconstructed from the raw telemetry, Fabricius et al. 2016): gaps in the spacecraft attitude solution deliberately introduced around the times when micro-meteoroid hits occurred (Lindegren et al. 2016). Telemetry losses along the spacecraft to ground link are only a very minor contribution to the data gaps. As a result of these gaps the amount of data processed for Gaia DR1 comprises slightly less than 12 (out of the above mentioned 14) months. The data gaps inevitably affect the quality of the Gaia DR1 results. In future releases the gaps related to the on-ground processing will disappear.

3. Overview of the contents of Gaia DR1

Gaia DR1 contains astrometry, G-band photometry, and a modest number of variable star light curves, for a total of 1 142 679 769 sources. Basic statistics for Gaia DR1 are listed in Table 1. The three main components of Gaia DR1 are:

1. The astrometric data set which consists of two subsets:
   - The primary astrometric data set contains the positions, parallaxes, and mean proper motions for 2 057 050 stars in common between the Gaia DR1, HIPPARCOS and Tycho-2 catalogues. This data set represents the realisation of the Tycho-Gaia astrometric solution (TGAS), of which the principles were outlined and demonstrated in Michalik et al. (2015). The typical uncertainty is about 0.3 mas for the positions, and about 1 mas yr⁻¹ for the proper motions. For the subset of 93 635 HIPPARCOS stars in the primary astrometric data set the proper motions are much more precise, at about 0.06 mas yr⁻¹. The typical uncertainty for the parallaxes is 0.3 mas, where it should be noted that a systematic component of ~0.3 mas should be added (see Sect. 6).
   - The secondary astrometric data set contains the positions for an additional 1 140 622 719 sources. For the secondary data set the typical uncertainty on the positions is ~10 mas. The positions and proper motions are given in a reference frame that is aligned with the International Celestial Reference Frame (ICRF) to better than 0.1 mas at epoch J2015.0, and non-rotating with respect to ICRF to within 0.03 mas yr⁻¹. The detailed description of the production of the astrometric solution, as well as a more detailed statistical summary of the astrometry contained in Gaia DR1 can be found in Lindegren et al. (2016). An in-depth discussion of the Gaia DR1 reference frame and the optical properties of ICRF sources is presented in Mignard et al. (2016).

2. The photometric data set contains the mean Gaia G-band magnitudes for all the sources contained in Gaia DR1. The brightest source in Gaia DR1 has a magnitude \( G = 3.2 \), while the majority of the sources (99.7%) are in the range \( 11.2 \leq G \leq 21 \). The small fraction of sources at \( G > 21 \) (where the nominal survey limit is \( G = 20.7 \), Gaia Collaboration 2016b) most likely have erroneously determined G-band fluxes, but nevertheless passed the data quality filters described in Sect. 4. The typical uncertainties quoted on the mean value of \( G \) range from a milli-magnitude or better at the bright end \( (G \leq 13) \), to about 0.03 mag at the survey limit. The details of the photometric data set, including the data processing and validation of the results is described in van Leeuwen et al. (2016), Carrasco et al. (2016), Riello et al. (2016), Evans et al. (2016).

3. The Cepheids and RR Lyrae data set contains the G-band light curves and characteristics of a modest sample of 599 Cepheid (43 newly discovered) and 2595 RR Lyrae (343 new) variables located around the south ecliptic pole and observed at high cadence during a special scanning period in the first four weeks of the operational phase of

<table>
<thead>
<tr>
<th>Source numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of sources</td>
</tr>
<tr>
<td>No. of primary (TGAS) sources</td>
</tr>
<tr>
<td>HIPPARCOS</td>
</tr>
<tr>
<td>Tycho-2 (excluding HIPPARCOS stars)</td>
</tr>
<tr>
<td>No. of secondary sources</td>
</tr>
<tr>
<td>No. of sources with light curves</td>
</tr>
<tr>
<td>Cepheids</td>
</tr>
<tr>
<td>RR Lyrae</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Magnitude distribution percentiles (G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.135%</td>
</tr>
<tr>
<td>2.275%</td>
</tr>
<tr>
<td>15.866%</td>
</tr>
<tr>
<td>50%</td>
</tr>
<tr>
<td>84.134%</td>
</tr>
<tr>
<td>97.725%</td>
</tr>
<tr>
<td>99.865%</td>
</tr>
</tbody>
</table>

Table 1. Basic statistics on the contents of Gaia DR1.
The magnitude distribution of the sources reveals a drop-off at $G \lesssim 7$. Neither all Hipparcos nor all Tycho-2 sources are included in Gaia DR1 and at the faint end the magnitude limit is sky position dependent and ill-defined. At magnitudes below $G \sim 5$ the total number of sources in Gaia DR1 is larger than the number of Hipparcos sources in Gaia DR1. This is however only apparent as most of these sources are in fact in common with the Hipparcos catalogue but have been treated as secondary astrometric sources, because a good 5-parameter astrometric solution could not be derived. The limitations of Gaia DR1, including its completeness, are discussed in Sect. 6.

Of the 1141 million sources in the secondary astrometric data set 685 million are in common with the Initial Gaia Source List (IGSL, Smart & Nicastro 2014) and 456 million are new sources (Lindegren et al. 2016). The IGSL formed the starting point for the process of assigning Gaia observations to sources (Fabricius et al. 2016). Hence the term “new” should strictly speaking be interpreted as referring to sources that could not be matched to known IGSL sources. No attempt was made to establish how many sources are truly new discoveries by Gaia but this is likely to be a substantial fraction (over 400 million) of the new sources mentioned above. The IGSL has been publicly available for some time and we caution that when looking up a source in Gaia DR1 through its already known IGSL source identifier, it should be kept in mind that a large fraction of the 1.2 billion sources in the IGSL does not appear in Gaia DR1.

4. **Gaia DR1 validation and source filtering**

A substantial effort was dedicated to the validation of the results contained in Gaia DR1. This is a complex task which takes place at various levels within the DPAC. The outputs produced by the DPAC subsystems (described in Gaia Collaboration 2016b) are validated first through an “internal” quality control process. For the astrometric data set in Gaia DR1 this internal validation is described in Lindegren et al. (2016), while that for the photometric and variable star data sets is described in Evans et al. (2016) and Eyer et al. (2016), respectively. A second validation stage is carried out by the DPAC unit responsible for the data publication (cf. Gaia Collaboration 2016b), which examines all the data contained in Gaia DR1 together and thus provides an independent quality check. This global validation process is described in Arenou et al. (2016). Here we summarise only the most important findings from the validation and provide complementary illustrations of the quality of Gaia DR1 in Sect. 5.

Numerous tests were done during the validation stage of the Gaia DR1 production, ranging from basic consistency checks on the data values to the verification that the data is scientifically correct. No problems were revealed that would prevent the timely publication of Gaia DR1. However, a number of minor problems were found that have been addressed either by a filtering of the available DPAC outputs before their incorporation into the data release, or by documenting the issues found as known limitations to Gaia DR1 (see Sect. 6). The filtering applied to the astrometric and photometric processing outputs before the global validation stage was as follows:

- For the primary astrometric data set only sources for which the standard uncertainties on the parallaxes and positions are less than 1 mas and 20 mas, respectively, were kept. In addition it was required that the sources have valid photometric data. For the secondary astrometric data set the sources were filtered by requiring that they were observed by Gaia at least 5 times (i.e. at least 5 focal plane transits), and that their astrometric excess noise (which indicates the astrometric modelling errors for a specific source) and position standard uncertainty are less than 20 mas and 100 mas, respectively. More details can be found in Lindegren et al. (2016). We stress that no filtering was done on the actual value of the source astrometric parameters.

- Although the photometric results were not explicitly filtered before their incorporation into Gaia DR1, a number of filters internal to the photometric data processing effectively leads to filtering at the source level. In particular sources with extremely blue or red colours will not appear in Gaia DR1.

- The only filtering done on the outputs of the variable star processing was to remove a handful of sources that were very likely a duplicate of some other source (see below for more discussion on duplicate sources).

The second validation stage (Arenou et al. 2016) revealed the following problems that were addressed through a further filtering of the astrometric and photometric processing outputs before their final incorporation into Gaia DR1. The filters described below were thus applied after the filters above.

- Some 37 million source pairs were found which are separated by less than 1 Gaia focal plane pixel size on the sky (i.e. 59 mas), or are separated by less than 5 times their combined positional standard uncertainty (where the factor 5 accounts for a possible underestimation of the standard uncertainties). The vast majority of these pairs are created during the cross-match stage, when observations (focal plane transits) get grouped together and assigned to sources (see Fabricius et al. 2016). The main underlying cause is sources appearing twice in the IGSL, which was evident from the many close pairs occurring along photographic survey plate boundaries (the IGSL is based to a large extent on photographic surveys, Smart & Nicastro 2014). A large fraction of these pairs are likely to be two instances of the same physical source (i.e. the source appears twice in the Gaia source list with two different identifiers). One member of each of these close pairs was filtered out of the Gaia DR1 source list and the remaining sources were flagged as having a duplicate associated to them in the Gaia source list. This flag thus
Fig. 2. Sky distribution of all Gaia DR1 sources in Galactic coordinates. The source density is shown with a grey scale chosen to highlight both the impressive amount of detail in the outlines of the well-known dust features along the Galactic plane, and the non-astronomical artefacts in the source distribution (see text). Image credits: CENTRA – University of Lisbon (part of the DPAC-CU9 visualisation team).

indicates that the source in question has fewer observations contributing to its astrometry and photometry because part of the observations were assigned to another (fictitious) source. This filtering will in a fraction of the cases inevitably have removed one component from a real double source (be it a binary or an optical pair). This problem of duplicate sources will disappear in future Gaia data releases due to improvements in the cross-match algorithm and the moving away from the Initial Gaia Source List as the basis for assigning observations to sources.

– For some 1 million sources the mean $G$ values were grossly inconsistent with either existing photometry (for example some TGAS stars were assigned $G$-band magnitudes much fainter than the Tycho-2 survey limit) or with the broad-band $G_{BP}$ and $G_{RP}$ magnitudes determined from the Gaia Blue and Red Photometers. In either case data processing problems are indicated and sources were removed from Gaia DR1 when there were fewer than 11 measurements in the $G$ band (i.e. CCD transits in the astrometric part of the focal plane), or if both ($G - G_{BP}$) and ($G - G_{RP}$) were larger than +3.

Although the filtering described above will have removed the vast majority of problem cases from the DPAC outputs before the publication of Gaia DR1, it will nevertheless not be perfect. Genuine sources will have been removed and the filtering criteria do not guarantee the absence of a small fraction of problematic sources in Gaia DR1.

The decision to filter out the problematic cases rather than publish them with, e.g. indicator flags, was driven by data quality considerations and by the need to remove the large number of spurious sources created in the process of matching observations to sources (see Fabricius et al. 2016; Lindegren et al. 2016). The filtering thus reflects the preliminary nature of the first Gaia data release. In future intermediate releases the shortcomings in the data processing will be addressed and more measurements will be added, which means that reliable results can be derived for more sources. The level of filtering is thus expected to go down and more sources will enter the published catalogue.

5. Illustrations of the Gaia DR1 contents

Here we provide a few illustrations of the contents of Gaia DR1. The purpose is not to provide a scientific analysis but to demonstrate through astronomically relevant examples the overall quality of the Gaia data. A more detailed examination of the scientific quality of Gaia DR1 is provided in two studies on open clusters (Gaia Collaboration 2016c) and the Cepheid period-luminosity relation (Gaia Collaboration 2016a). We end this section with a comment on the Pleiades cluster distance.

5.1. The Gaia sky

The distribution of all Gaia DR1 sources on the sky is illustrated in Fig. 2. The source density shown in Fig. 2 is based on the accurate positions of the 1.1 billion sources in Gaia DR1 and represents the most detailed all-sky map in the optical to date. This can be appreciated in particular in the very fine outlining of the dust features along the Galactic plane. Also noteworthy are the Magellanic Clouds, where in the Large Magellanic Cloud the individual features in the star forming regions north of the bar are outlined in the source distribution; the M 31 and M 33 galaxies which are both outlined in individual detections made by Gaia; and the Orion A and B clouds which can be seen against the backdrop of the sources detected by Gaia. Also recognisable are globular clusters, such as ω Centauri with over two hundred thousand sources individually appearing in Gaia DR1, and the Fornax dwarf spheroidal galaxy (∼30 000 sources in Gaia DR1) near $(\ell, b) \approx (237^\circ, -66^\circ)$. The full detail of this sky map is impossible to convey in print. An interactive and zoomable version will be available, through the Aladin sky atlas application (Bonnarel et al. 2000; Boch & Fernique 2014) and a dedicated visualisation service, both as part of the Gaia
The depth of the Gaia survey, its all-sky reach, the high angular resolution, combined with the highly accurate source positions, promises a revival of classical star count studies, in particular with future Gaia data releases where the shortcomings in the completeness and angular resolution of Gaia DR1 (see Sect. 6) will have been addressed. The Gaia sky map is also of immediate interest to studies of minor solar system bodies through stellar occultations, the predictions of occultation tracks on the Earth benefiting from the dense distribution of sources with accurately known positions.

Finally, the Gaia sky map will be the standard reference in the optical for some time to come, in particular when in future releases the Gaia catalogue will be more complete in sky, magnitude, and colour coverage, and the source positions are further refined, with parallaxes and proper motions becoming available for all Gaia sources. This is to the benefit of all (optical) telescope guidance applications, especially large-mirror telescopes with small fields of view. Space missions will also benefit from the Gaia sky map. As an example, it is planned to improve the recently released Hubble Source Catalog (Whitmore et al. 2016) through a re-reduction of the astrometry with respect to the Gaia source positions.

5.2. Hertzsprung-Russell diagrams based on Gaia DR1 parallaxes

With the advent of Gaia DR1 we now for the first time have access to two large samples of parallaxes accurate at the (sub-) milliarcsecond level. As explained in Lindegren et al. (2016) the Gaia and HIPPARCOS parallaxes are independent and can thus sensibly be compared to each other. The comparison described in the appendix of Lindegren et al. (2016) shows that overall the Gaia DR1 and HIPPARCOS parallaxes are the same to within the combined uncertainties. A closer look at the parallaxes near zero reveals that for the HIPPARCOS stars in Gaia DR1 the number of negative parallaxes is much smaller, which is expected for a data set that is more precise. This comparison is furthermore exploited in Lindegren et al. (2016) to derive the relation between the formal and actual (published) uncertainties for the astrometric source parameters in the primary astrometric data set of Gaia DR1.

We illustrate the better overall precision of the Gaia parallaxes by constructing observational Hertzsprung-Russell (HR) diagrams in $M_G$ vs. $(B-V)$ using the HIPPARCOS parallaxes from van Leeuwen (2007) and the parallaxes from Gaia DR1. The result is shown in Fig. 3. The 43 546 HIPPARCOS stars included in the left two panels a and b were selected according to:

\[
(\sigma_{\pi}/\pi)_{Gaia} \geq 5 \land (\sigma_{\pi}/\pi)_{HIPPARCOS} \geq 5 \land \\
\sigma_{\pi} \leq 0.05 \land \sigma_{(B-V)} \leq 0.05,
\]

where $\pi$ is the parallax and $\sigma_{\pi}$ the corresponding standard uncertainty. The values of $(B-V)$ and their standard uncertainties were taken from the HIPPARCOS Catalogue (van Leeuwen 2007). The 74 771 stars in the rightmost panel (c) were selected only on the value of the relative uncertainty in the Gaia DR1 parallax but with the same criteria on the uncertainty in $G$ and $(B-V)$. The median Gaia DR1 parallax for the smaller sample is 7.5 and for the larger sample it is 5.0 mas, while 90 per cent of the stars have a parallax larger than 3.6 (smaller sample) and 2.2 mas (larger sample). A comparison of the left two panels shows that with the Gaia DR1 parallaxes the main sequence is better defined, being somewhat narrower and with a sharper boundary along the faint end. The distribution of red clump giants is much narrower in luminosity, with the effect of extinction and reddening clearly seen as an elongation in the direction of fainter magnitudes and redder colours.

The narrower luminosity distribution of the red clump giants and main sequence dwarfs in Gaia DR1 is further illustrated in Fig. 4. The luminosity distribution is shown for the stars in the left two panels of Fig. 3 that have colours in the range $1.0 \leq (B-V) \leq 1.1$ (3174 stars), including both the clump stars...
around $M_G \sim 0.5$ and the main sequence dwarfs around $M_G \sim 6$, as well as a fraction of sub-giants (at $1 \leq M_G \leq 3$). The luminosity distributions for both the dwarfs and the clump giants are narrower for Gaia DR1 than for HIPPARCOS. For the dwarfs (defined as stars with $M_G > 4.5$) the robust scatter estimates for the width of the distribution of $M_G$ (see Lindegren et al. 2016, for the definition of this quantity) are 0.32 for Gaia DR1 and 0.38 for HIPPARCOS. For the clump giants the numbers are sensitive to the range of $M_G$ used to isolate the clump and whether that range is defined using the Gaia DR1 or HIPPARCOS luminosities. Using the broad selection $-0.5 \leq M_G(\text{HIPPARCOS}) \leq 1.5$ the robust scatter estimates are 0.37 for Gaia DR1 and 0.46 for HIPPARCOS. When isolating the clump giants using Gaia DR1 luminosities ($-0.2 \leq M_G(\text{Gaia DR1}) \leq 1.2$) the robust scatter estimates are 0.30 for Gaia DR1 and 0.49 for HIPPARCOS. The detailed interpretation of the scatter in $M_G$ for the red clump giants and how this relates to the parallax quality of Gaia DR1 and HIPPARCOS is complicated by the HIPPARCOS survey selection function, the filtering applied for Gaia DR1, the parallax systematics present in Gaia DR1 (see Sect. 6), and the biases introduced by the HIPPARCOS magnitude limit and the selection on relative parallax error. These effects lead to an incomplete and non-representative sample of red clump giants. A proper interpretation of the scatter in the luminosities (and of the mean observed luminosity) requires the modelling of the population of red clump giants and of the Gaia and HIPPARCOS survey properties, which we consider beyond the scope of the illustrations provided in this section.

The rightmost panel in Fig. 3 shows how in Gaia DR1 a larger volume is covered by relatively precise parallaxes; the overall width in colour of the upper main sequence and red clump is larger due to the larger extinction values probed, and the upper main sequence and giant branch are better populated. In numbers the median relative uncertainty on the HIPPARCOS parallax for the stars selected according to Eq. (2) is 0.1, while for the Gaia DR1 parallaxes it is 0.04.

In Fig. 5 we show the observational HR diagram for a much larger sample of stars from Gaia DR1 for which the $(G - K_s)$ colour index can be calculated from Gaia DR1 and the data in the 2MASS Point Source Catalogue (Skrutskie et al. 2006). The selection of the sources in this diagram is according to Eq. (2), where the limit on the HIPPARCOS relative parallax error does not apply and the limit on the standard uncertainty in the colour index now applies to $(G - K_s)$. In addition the 2MASS photometric quality flag was required to be equal to “A” for each of the $J$, $H$, and $K_s$ magnitudes. The resulting sample contains 1,004,204 stars (there are 1,037,080 stars with $\varpi/\sigma_\varpi \geq 5$ in total in Gaia DR1). The sample covers a substantially larger volume, the median parallax being 2.9 mas, while 90 per cent of the stars have a parallax larger than 1.7 mas. The larger volume covered is evident from the large number of luminous stars in the HR diagram: 42,333 stars at $M_G < 2$ in the rightmost panel of Fig. 3, compared to 190,764 in Fig. 5. In addition the effect of extinction is now more prominently visible as a broadened upper main sequence and turn-off region, as well as in the elongation of the red clump. A hint of the binary sequence in parallel to the main sequence can be seen around $(G - K_s) \sim 2.2$ and $M_G \sim 6$. Note the three white dwarfs at $(G - K_s) < 0$ and $M_G > 11$; from the brightest to the faintest their 2MASS designations and Gaia source identifiers are 2MASS J21185627+5412413, 2MASS J16482562+5903228, 2MASS J19203492-0739597, and 2176116580055936512, 1431783457574556672, 42017812704230716, respectively. This diagram also has the purpose of the pre-computed cross-match tables, linking Gaia DR1 and other large surveys, which are provided along with the data release (see Sect. 7 and Marrrese et al. 2016).

An HR diagram can also be produced with the $(B - V)$ colour index. However, this requires the use of different sources for the colour index values. When we combined HIPPARCOS
All-sky optical photometry over the brightness range in between third as many stars resulted. This reflects the lack of high quality APASS (Henden & Munari 2014) a diagram containing only a only cover apparent magnitudes fainter than \( \sim \) as the Sloan Digital Sky Survey (York et al. 2000), which usually (ESA 1997; van Leeuwen 2007), Tycho-2 (Høg et al. 2000), and APASS (Henden & Munari 2014) a diagram containing only a third as many stars resulted. This reflects the lack of high quality all-sky optical photometry over the brightness range in between that covered by HIPPARCOS and modern digital sky surveys, such as the Sloan Digital Sky Survey (York et al. 2000), which usually only cover apparent magnitudes fainter than \(-15\). This situation will be remedied with the second Gaia data release through the publication of the \( G_{BP} \) and \( G_{RP} \) magnitudes obtained from the integrated fluxes measured with the Blue and Red Photometers.

Finally, following Gould (2004), in Fig. 6 we show a version of the HR diagram which is colour coded according to the transverse velocity of the stars \( v_\perp = \mu/\sigma \times 4.74 \ldots \) (in km s\(^{-1}\)), where \( \mu \) is the length of the proper motion vector of the star. The stars in this diagram were selected according to the criteria in Eq. (2) (without selecting on the HIPPARCOS relative parallax error), using the \((B - V)\) colour index as listed in the HIPPARCOS, Tycho-2, or APASS catalogues (in that order of preference), where the Tycho-2 colours were transformed to approximate Johnson colours according to: \((B - V)_J \approx 0.85(B - V)_T\) (ESA 1997, Sect. 1.3, Vol. 1). It was further demanded that \( G \leq 7.5 \), or \( \mu \geq 200 \) mas yr\(^{-1}\), or \( \sigma \geq 10 \) mas. The 41 136 stars in this diagram are represented by symbols which are colour coded by tangential velocity interval as indicated in the figure legend. This nicely illustrates the well-known mix of stellar populations in a local sample (the median parallax for this sample being 10.7 mas, while 90 per cent of the stars have a parallax larger than 2.8 mas). At low velocities the young disk stars along the main sequence are outlined \( (v_\perp < 50 \) km s\(^{-1}\)). The turn-off region for the old disk is visible at \( 50 \) km s\(^{-1}\) \( \leq v_\perp < 100 \) km s\(^{-1}\), while at higher velocities halo stars are visible, which along the main sequence are clearly shifted to the lower metallicity region.

5.3. Gaia DR1 proper motions

Given the different time spans that underlie the determinations of proper motions listed in the HIPPARCOS (\( \Delta \)epoch \( \sim \) 3.5 yr), Gaia DR1 (\( \Delta \)epoch \( \sim \) 24 yr), and Tycho-2 (\( \Delta \)epoch \( \sim \) 90 yr) catalogues, it is interesting to look for sources with discrepancies between the proper motions listed in the three catalogues. The proper motion differences may point to the presence of non-modelled astrometric components (such as orbital motion in a binary), and thus to sources worthy of further investigation.

If this is attempted, very large discrepancies between Gaia DR1 and Tycho-2 proper motions may occur (of order \( 100 - 250 \) mas yr\(^{-1}\)), which seems surprising at first sight. We performed a close inspection of 39 such cases and examined proper motion solutions for these sources for which the Tycho-2 position was not used (these solutions are not published in Gaia DR1). In all cases there is close agreement (to within a few mas yr\(^{-1}\) in both coordinates) between the Gaia-only proper motion and the proper motion listed in Gaia DR1. The fact that Gaia measures the same proper motions over a 14 month time span as over the 24 yr time span used for the primary astrometric data set implies that the large discrepancies mentioned above are due to errors in the Tycho-2 proper motions. These errors are most likely caused by mismatches of the Tycho sources to old photographic catalogues, as was confirmed by inspecting the surroundings of a few sources among the 39 mentioned above.

The above example points to the high quality of the Gaia DR1 proper motions and serves as a warning not to over-interpret discrepancies between Gaia DR1 proper motions and those in existing proper motion catalogues.

5.4. Photometry of variable stars

Figure 7 shows two examples of phase-folded light curves from the Cepheids and RR Lyrae data set in Gaia DR1, one of a Cepheid and one of an RR Lyrae variable. Both curves highlight the quality of the \( G\)-band photometry in Gaia DR1. In the case of the Cepheid variable the error bars are comparable to or smaller than the symbol size, while for the RR Lyrae variable the uncertainties on the individual measurements are \( \sim 0.02 \) mag. More light curves and an extensive description of the Cepheids and RR Lyrae variables in Gaia DR1 are presented in Clementini et al. (2016). The high cadence at which these stars were observed is not representative for the nominal Gaia mission, but reflects the special Ecliptic Pole Scanning Law used during the first weeks of the mission (Gaia Collaboration 2016b).

5.5. Comment on the Pleiades cluster mean parallax

Since the publication of HIPPARCOS-derived trigonometric cluster parallaxes for the Pleiades (van Leeuwen 1999, 2009) there has been a discrepancy between the HIPPARCOS values and a number of other distance determinations derived with various methods. Figure 8 displays the set of existing measurements of either the parallax or the distance modulus of the cluster or of individual cluster members, all expressed as distances in parsecs. The Gaia DR1 adds another item to this set. It is indicated in Fig. 8 by the yellow shaded area.

Fig. 6. Observational HR diagram showing where stars with specific values of the transverse velocity \( v_\perp \) tend to occur. The colour coding of the points is according to tangential velocity interval, as indicated in the legend (in km s\(^{-1}\)).
Fig. 7. Example light curves from the Cepheids and RR Lyrae data set in Gaia DR1. The top panel shows the light curve for a fundamental mode classical Cepheid in the Large Magellanic Cloud (period 2.891 days), while the bottom panel shows the light curve for a fundamental mode RR Lyrae star (RRab, period 0.607 days), also in the Large Magellanic Cloud.

A simplistic selection of Pleiades members can be done solely on the basis of the Gaia DR1 positions and proper motions by demanding that the selected stars lie within 5 degrees from the position \((\alpha, \delta) = (56.75^\circ, 24.12^\circ)\) and that the proper motions obey:

\[
\left( \mu_\alpha \cos \delta - 20.5 \right)^2 + \left( \mu_\delta + 45.5 \right)^2 \leq 6 \text{ mas yr}^{-1}.
\]

This leads to the selection of 164 stars from the Gaia DR1 primary astrometric data set. Figure 9 shows the histogram of the parallaxes of these 164 stars, which apart from a few outliers (field stars not belonging to the Pleiades) are well clustered in a peaked distribution. The median of this distribution is at \(\varpi = 7.45\) mas, and the standard deviation (robustly estimated) of the distribution is 0.49 mas. If the observations were independent, this would lead to a standard uncertainty in the mean of 0.49/\(\sqrt{N}\) = 0.04 mas. However, as described in the paper on the astrometric solution for Gaia DR1 (Lindegren et al. 2016) and in the paper on the validation of Gaia DR1 (Arenou et al. 2016), a not precisely known systematic uncertainty of the order of 0.3 mas must be added to the parallax uncertainties (see also Sect. 6). These systematic terms are correlated over small spatial scales, which means that the parallax uncertainties are not independent for the Pleiades members considered here, leading to no reduction of the uncertainties by averaging. Therefore the best estimate we can make at this time for the mean Pleiades parallax is 7.45 ± 0.3 mas, corresponding to a distance of about 134 ± 6 pc. This is indicated by the half-width of the yellow shaded area in Fig. 8.

We want to emphasise that, taking this systematic uncertainty into account, Gaia DR1 cannot be considered as giving a final and definite answer on the so-called Pleiades distance discrepancy. In particular an explanation for the discrepancy between Gaia DR1 and HIPPARCOS cannot be provided at this stage. A proper and more extensive analysis of the Gaia DR1 astrometry for nearby open clusters (including the Pleiades) is presented in Gaia Collaboration (2016c), with the results providing further arguments as to why the Pleiades distance estimated from Gaia DR1 parallaxes cannot be considered definitive.
conclusive answer to the question on the Pleiades distance – in the form of a sufficiently precise and systematically reliable trigonometric parallax for the cluster – can, however, be expected from future Gaia data releases (probably already Gaia DR2). What the present release definitely does is to make another significant addition to the accumulating information on the Pleiades distance which is summarised in Fig. 8.

6. Known limitations of Gaia DR1

Gaia DR1 represents a major advance in terms of the availability of high-accuracy parallaxes and proper motions for the 2 million stars in the primary astrometric data set and in terms of accurate positions and homogeneous all-sky photometry for all sources out to the Gaia survey limit. Nevertheless the data release is based on an incomplete reduction of a limited amount of raw Gaia data and is thus of a very preliminary nature. We summarise the major shortcomings of Gaia DR1 in this section both to warn the users of the data and to enable a careful scientific exploitation of the Gaia DR1 data set. We stress however, that all the shortcomings listed below will be addressed in future Gaia data releases, with major improvements already expected for the second data release.

6.1. Data processing simplifications for Gaia DR1

We show in Fig. 10 in highly simplified form the DPAC data processing flow for the astrometric and photometric data reduction. The purpose of the diagram is to highlight the shortcomings in the data processing for Gaia DR1 compared to the intended data processing for future data releases (for simplicity many processing steps are left out, including the processing of the RVS data and the derivation of higher level results, such as source astrophysical parameters). The steps that should be taken during the processing are:

1. From the raw data derive (initial) calibrations of the Gaia PSF, the CCD bias, the astrophysical and stray light induced background flux in the image, and the parameters describing the charge transfer inefficiency (CTI) effects in the CCDs.
2. Use the calibrations to determine from the raw CCD-level measurements both the source flux and the source location within the observation window.
3. Use the spacecraft attitude to create the source list, by assigning observations (focal plane transits) to existing sources or by creating new sources if needed.
4. Process the image fluxes to derive calibrated G-band photometry and process the BP/RP data to derive the source colours. Process the image locations in order to derive the astrometric source parameters, the attitude model for the Gaia spacecraft, and the geometric instrument calibrations.
5. Introduce the known source locations on the sky, the geometric instrument calibrations, the attitude model, and the source colours into step 1 above and improve the accuracy of the calibrations.
6. Repeat steps 2 and 3 using the improved astrometry and calibrations from step 5. Subsequently repeat step 4 using the improved image locations and fluxes.
7. Iterate the above steps, including progressively more data, until convergence on the final astrometric and photometric results at their ultimately attainable accuracy.

As illustrated in Fig. 10 steps 5–7 above were not carried out during the processing for Gaia DR1, which means that the inputs for the astrometric and photometric processing are limited in quality due to the use of immature calibrations, in particular an incomplete PSF model which does not account for source colour effects on the detailed image shape, or for PSF variations across the focal plane and in time. The source locations within the images and the astrometry derived from those will be strongly affected by systematics related to source colour (see Lindegren et al. 2016, appendix C). Systematic effects related to the PSF model can also be expected in the G-band photometry derived from the image fluxes. A further limitation to the quality of Gaia DR1 astrometry, indicated in Fig. 10, is that the attitude modelling within the astrometric solution is incomplete. No treatment of micro-meteoroid hits or micro-clanks was included (except for the exclusion of the data from short time intervals affected by large hits) leading to attitude modelling errors which in turn will limit the astrometric accuracy that can be attained (see Lindegren et al. 2016, in particular appendix D). The treatment of CTI effects was not included in Gaia DR1, which is justified given the present low levels of radiation damage to the Gaia CCDs (Crowley et al. 2016a).

We stress that the above description of the data processing for Gaia DR1 is mainly illustrative and not intended as a complete description of all the simplifications that were introduced to enable a timely first Gaia data release. For details on the actual processing for Gaia DR1 refer to Fabricius et al. (2016) (pre-processing and source list creation), van Leeuwen et al. (2016), Carrasco et al. (2016), Riello et al. (2016) (photometric processing), Eyer et al. (2016) (variable star processing), and Lindegren et al. (2016) (astrometric processing). In particular the latter paper contains an extensive description of the known problems introduced by the preliminary astrometric processing.

In the following subsections we summarise the most prominent issues with Gaia DR1 which should be taken into consideration when using the data for scientific analyses. These concern catalogue completeness, and systematics in the astrometric and photometric results which were revealed during the validation of the DPAC outputs produced for Gaia DR1. Much more detail on the validation of Gaia DR1 can be found in
6.2. Gaia DR1 source list and completeness

The Gaia DR1 celestial source density distribution shown in Fig. 2 contains a number of clearly non-astronomical artefacts, which is illustrated in more detail in Fig. 11 for the Milky Way bulge region. In particular away from the Milky Way plane, but also across the Bulge region, Fig. 2 shows obvious source underdensities as well as apparent over-densities, where the latter surround the regions (along the ecliptic) dominated by the former.

The patterns in Fig. 2 are related to the Gaia scanning law (cf. Gaia Collaboration 2016b) and are caused by the source filtering applied for Gaia DR1. The areas around the ecliptic are inherently observed less often due to the characteristics of the scanning law, and in particular have been rather poorly observed over the first 14 months of the mission (covering Gaia DR1), both in terms of the number of visits and the coverage in scanning direction. This results in the sources in the less well covered areas having a larger probability of being filtered out, which gives the regions in between (with far fewer sources filtered out) the appearance of containing more sources. Hence Fig. 2 shows primarily a deficit of sources in the less well observed regions of the sky.

This is illustrated in more detail for the Milky Way bulge region in Fig. 11. The pattern of dark stripes, with a clear lack of sources, is again related to the Gaia scanning law. The bulge lies in the ecliptic region and thus suffers from poor scan law coverage in Gaia DR1. In combination with the filtering on the astrometric solution quality prior to Gaia DR1 publication this can even lead to areas where sources are entirely missing. This is illustrated in Fig. 12, which shows the circle on the sky around $(\alpha, \delta) = (266^\circ, -18.5^\circ)$ with a 0.5 degree radius. The top panel shows the distribution of the 268435 sources in this area. The distribution shows the striping pattern and also contains very thin strips where fewer sources than the average are found. Most prominent, however, are three large gaps where no
sources occur. The middle panel shows the median number of observations matched to each source in this region and there the pattern is even richer. We note that the minimum number of observations is five (as demanded by the filtering applied, see Sect. 4), suggesting that the gaps are related to the number of times a particular coordinate on the sky was visited by Gaia. This is confirmed in the bottom panel which shows a simulation of the expected number of visits corresponding to the scanning law as executed between September 2014 and September 2015. This time period does not cover the ecliptic pole scanning phase, but during that phase this region on the sky was not observed. The gaps in the source distribution correspond to the areas in the simulation where fewer than five visits by Gaia occur, which thus explains the gaps as being due to the filtering applied for Gaia DR1. In addition the simulation shows the same very thin strips where Gaia has collected fewer observations than the maximum of 12 occurring in this area on the sky. Whenever few observations are collected there is a good chance that the source gets filtered out if focal plane transits from particular visits by Gaia are discarded for other reasons and thus the total number of observations drops below five. Although the simulated scan law coverage very much resembles the pattern in the number of matched observations, there are differences in detail because the actually executed scanning law differs somewhat from the nominal scanning law used in the simulation.

The stripping pattern seen over the bulge region in Fig. 11 can thus be explained as a consequence of the Gaia scan law coverage over the first 14 months of the mission combined with the filtering applied to the astrometric results before including them in Gaia DR1. Although the stripping and gaps are most prominent visible in the bulge region this pattern also occurs in other parts of the sky in the ecliptic region, notably along the Milky Way plane in the anti-centre direction. In these areas the step changes in the number of observations collected by Gaia combined with the filtering has in some unlucky cases led to one half of an open cluster partly missing from the catalogue.

Further remarks on the catalogue completeness are the following:

- Many bright stars at $G \lesssim 7$ are missing from Gaia DR1 as the corresponding measurements cannot yet be treated routinely by the DPAC. The images are heavily saturated and the instrument configuration (TDI gate setting used) is difficult to calibrate due to the sparsity of bright sources on the sky.
- High proper motion stars ($\mu > 3.5$ arcsec yr$^{-1}$) are missing from the catalogue due to a technical issue in the construction of the IGSL (cf. Lindegren et al. 2016).
- As mentioned in Sect. 4 extremely blue and red sources are missing from Gaia DR1 which, for example, affects the completeness of the white dwarf population in Gaia DR1 and that of sources in extincted regions (cf. Arenou et al. 2016).
- In dense areas on the sky (with source densities above a few hundred thousand per square degree) the crowding of sources will lead to the truncation of the observation windows for some stars when they overlap with the window of another star. These truncated windows have not been used in the data processing for Gaia DR1. This means that in dense areas the average number of transits used per source will be smaller (especially for fainter sources), which in combination with the filtering on the number of observations and the astrometric or photometric solution quality means these sources may have been removed from Gaia DR1.
- The survey completeness is also affected by the way the data is treated on board Gaia, meaning both the detection of sources and the assignment of observation windows. The details are provided in Gaia Collaboration (2016b). We note here that in very dense areas (above $\sim 400000$ stars per square degree) the effective magnitude limit of the Gaia survey may be brighter by up to several magnitudes, with data for faint sources being collected for a reduced number of focal plane transits.

- An examination of double stars from the Washington Visual Double Star Catalog (Mason et al. 2001) contained in Gaia DR1 shows that below about 4 arcsec there is a notable decrease in the completeness of the detection of the secondaries, which is related to the above mentioned limitations in crowded regions (Arenou et al. 2016). The implication of this finding and the previous two items is that the effective angular resolution on the sky of Gaia DR1, in particular in dense areas, is not yet at the levels expected for the 1.5 m Gaia telescope mirrors (which should lead to an angular resolution comparable to that of the Hubble Space Telescope).

The limitations to the Gaia DR1 source list described above lead to a catalogue which is not complete in any sense and for which the faint magnitude limit is ill-defined and dependent on celestial position. No attempt was made to derive a detailed completeness function. Hence when using the catalogue for scientific analyses, care needs to be taken with the interpretation of source distributions both on the sky and in apparent magnitude.

6.3. Known problems in the Gaia DR1 photometry

Although the $G$-band fluxes and magnitudes provided with Gaia DR1 have standard uncertainties as good as a few per cent in magnitude at the survey limit and down to the milli-magnitude level at the bright end, there are nevertheless limitations inherent to this first Gaia data release. The $G$-band fluxes were derived as part of the image parameter determination in the initial data treatment (see Sect. 6.1 above and Fabricius et al. 2016) and thus suffer from the lack of an accurate PSF model. In addition at the bright end ($G < 12$) the calibrations of the photometry are complicated by the use of TDI gates, while over the range $G = 12–17$ the effects of different observation window sizes make the calibration more complex. The result is that for the brightest, $G < 12$, stars the photometric accuracy is estimated to currently be limited to a calibration floor of $\sim 3$ mmag for the individual CCD transits (van Leeuwen et al. 2016; Evans et al. 2016). The quoted standard uncertainties on the mean $G$-band magnitudes at the bright end can vary by an order of magnitude (caused by poorly calibrated transitions from one TDI gate setting to another). Over the range $G = 12–17$ the distribution of photometric standard errors as a function of magnitude shows two bumps at $G \sim 13$ and $G \sim 16$ which are related to the transition from one observation window type to another (van Leeuwen et al. 2016). An examination of the scatter in repeated photometric measurements for well-observed sources indicates that the quoted standard uncertainties on the $G$-band photometry are largely realistic as indicators of the photometric precision (see Evans et al. 2016, for details), however unaccounted for systematic errors cannot be excluded. Potential systematic errors in the photometry are discussed in Evans et al. (2016) and Arenou et al. (2016). There is a small fraction of sources for which the mean value of $G$ is clearly wrong. These are sources with magnitudes well beyond the Gaia survey limit of $G = 20.7$ and also at brighter magnitudes such errors occur as evidenced by the presence of a small number of Tycho-2 sources with magnitudes up to $G \sim 19$ (cf. Fig. 13), although it should be
noted that a number of these sources may well be variables with large brightness excursions, leading to faint magnitudes at the Gaia DR1 observation epoch.

6.4. Known problems in the Gaia DR1 astrometry

The data processing shortcuts and simplifications discussed in Sect. 6.1 have introduced a number of known weaknesses in the astrometric solution for Gaia DR1, which are described and explained extensively in Lindegren et al. (2016). Here we highlight the weaknesses most directly relevant to the scientific exploitation of the Gaia DR1 data.

Source modelling. All sources were treated as single stars without taking their radial velocity into account. Hence any astrometric effects due to the orbital motion in binaries or due to perspective acceleration were ignored. In addition for resolved binaries the positions used to derive the mean proper motion over the time period between the HIPPARCOS/Tycho (around J1991.25) and the Gaia DR1 (J2015.0) epochs may be inconsistent (cf. Lindegren et al. 2016). The Gaia DR1 catalogue provides the so-called excess source noise, which is meant to represent the astrometric modelling errors for a specific source, and thus could in principle be used to identify candidate astrometric binaries or otherwise problematic sources. However in Gaia DR1 all sources have significant excess source noise because currently unmodelled attitude and calibration errors are partly “absorbed” in this quantity (see Lindegren et al. 2016, for more details). The level at which the excess source noise is indicative of a source being different from a single star should thus be calibrated against a sample of known non-single star sources in Gaia DR1 before it can be used in scientific analyses.

Periodic basic angle variations. As described in Gaia Collaboration (2016b), a number of issues affecting the performance of the Gaia instruments came to light during the commissioning period. The most relevant issue for the astrometric quality of Gaia DR1 is the periodic variation of the basic angle between the two telescopes of Gaia. This angle enters into all the measurements of angular separations between sources on the sky and its value should either be stable or its variations known at the level of ~1 μas. The actual basic angle variations, measured both through the on board metrology system and from the daily astrometric solution carried out as part of the DPAC First-Look analysis (see Fabricius et al. 2016), have a component which varies periodically with the satellite spin period and with a significant amplitude of roughly 1 mas. The harmonic component that varies as the cosine of the spacecraft heliotropic spin phase cannot be distinguished from a zero-point offset in the parallaxes, making the calibrations of the basic angle variations an essential component of the success of Gaia (for more detail see Michalik & Lindegren 2016). For Gaia DR1 the corrections for the basic angle variations were done by adopting the variations as measured by the on board metrology system. At the accuracy level of Gaia DR1 this is sufficient. However Lindegren et al. (2016) do conclude that a global parallax zero point offset of ±0.1 mas may be present, which is confirmed by the zero-point offset of about −0.04 mas found during the validation of Gaia DR1 (Arenou et al. 2016). For future data releases the basic angle variations will largely be determined as calibration parameters within the astrometric solution (cf. Lindegren et al. 2016) with the aim to fully account for the variations.

Strongly correlated astrometric parameters. Figure 7 in Lindegren et al. (2016) presents a statistical overview of the standard uncertainties and the correlations between the astrometric parameters of each source in the primary astrometric data set. In Gaia DR1 the correlation levels are high, reaching median values near ~1 or +1 over large regions of the sky. It is thus very important to make use of the full covariance matrix when taking the standard uncertainties on (subsets and linear combinations of) the astrometric parameters into account in any scientific analysis of the data. The correlations will decrease in future data releases as the number of observations per source and the scan direction diversity increase.
Spatially correlated systematics. Several of the weaknesses in the astrometric solution identified in Lindegren et al. (2016) will lead to systematic errors that are colour dependent and spatially correlated over areas on the sky that may extend up to tens of degrees. One important contributor to these correlations is the incomplete modelling of the spacecraft attitude, which is extensively described in appendix D.3 of Lindegren et al. (2016). Special astrometric validation solutions indeed point to the presence of spatially correlated and colour-dependent systematics of $\pm 0.2$ mas. The global validation of the astrometric results confirms the presence of spatial variations of the parallax zero-point (see Arenou et al. 2016). Over large spatial scales the parallax zero-point variations reach an amplitude of $\sim 0.3$ mas, while over a few smaller areas ($\sim 2$ degree radius) much larger parallaxes biases may occur of up to $\pm 1$ mas. The recommendation is to consider the quoted uncertainties on the parallaxes as $\sigma \pm \sigma_{\sigma}$ (random) $\pm 0.3$ mas (systematic). Furthermore averaging parallaxes over small regions of the sky will not reduce the uncertainties to less than 3 mas level. Systematicities into proper motion biases are not possible due to the limited accuracy of ground-based proper motion catalogues.

Finally, we illustrate graphically the preliminary nature of the Gaia DR1 astrometry in Fig. 13. It shows the distribution of parallax standard uncertainties as a function of magnitude for HIPPARCONS (van Leeuwen 2007) and Gaia DR1, and the expected parallax standard uncertainties achievable after a 5-yr Gaia mission (as provided in Gaia Collaboration 2016b). Note how in contrast to HIPPARCONS the Gaia DR1 parallax standard uncertainties do not decrease with increasing source brightness but stay at the same level. This is partly related to the gating strategy for bright ($G < 12$) sources which prevents significant gains in signal to noise ratio, but the uncertainty levels are much more than a factor of 2 away from the 5-year mission uncertainty floor at the bright end (where the factor of 2 is the gain in signal to noise going from 14 to 60 months of observations). This indicates that the parallax uncertainties are dominated by calibration errors at this stage, the calibration floor being $\sim 0.2$ mas. The second important point in this figure is that the expected 5-yr parallax standard uncertainties are much better than what can be achieved for Gaia DR1, with the current parallax standard uncertainty levels being comparable to the standard uncertainty levels that can ultimately be achieved at the Gaia survey limit.

7. Gaia DR1 access facilities

Access to the data contained in Gaia DR1 is provided through various channels. The main access point is the ESA Gaia Archive, which can be accessed through http://archives.esac.esa.int/gaia/. The archive provides access to the data through simple query forms but also allows the submission of sophisticated data base queries in the Astronomical Data Query Language (Osuna et al. 2008). The electronic tables comprising Gaia DR1 contain descriptions of each data field which can be inspected online. The Gaia archive is Virtual Observatory (http://www.ivoa.net/) compatible and also allows for access through the Table Access Protocol (Dowler et al. 2010). More extensive documentation, providing more detail on the data processing than is possible to include in peer-reviewed papers, is available from the archive in various electronic formats. Further tools provided are a visualisation application, graphics with statistical overviews of the data, an online help system, and the means to upload user generated tables which can be combined with Gaia data and shared with other users of the Gaia archive. More details on the data access facilities are provided in Salgado et al. (2016).

As part of the archive services pre-computed cross-match tables linking Gaia DR1 to other large surveys are provided to facilitate the analysis of combined data sets. The details on how these cross-match tables were computed are provided in Marrese et al. (2016).

Finally, the Gaia DR1 data is also made available through a number of partner and affiliated data centres located in Europe, the United States, South Africa, and Japan. These data centres do not necessarily hold all the data contained in the Gaia archive and may layer their own access and analysis facilities on top of the Gaia data.

8. Conclusions

Less than three years after the launch of Gaia we present the first Gaia data release, where the use of positional information from the HIPPARCOS and Tycho-2 catalogues allowed the derivation of positions, parallaxes, and proper motions for about 2 million sources from the first 14 months of observations. This represents a data release that was not foreseen in the original Gaia mission planning and presents the astronomical community with advanced access to a large set of parallaxes and proper motions for sources to magnitude 11.5, at precisions substantially better than previously available. The release contains the positions and the mean $G$-band magnitudes for an additional 1141 million sources to the Gaia survey limit at $G \approx 20.7$, as well as the light curves for a sample of about three thousand variable stars.

The typical uncertainty for the position and parallaxes for sources in the primary astrometric data set is about 0.3 mas, and about 1 mas yr$^{-1}$ for the proper motions. We stress again that a systematic component of $\sim 0.3$ mas should be added to the parallax uncertainties and that averaging parallaxes over small regions on the sky will not lead to a gain in precision. For the subset of HIPPARCOS stars in the primary astrometric data set the proper motions are much more precise, at about 0.06 mas yr$^{-1}$ (albeit with a systematic uncertainty at the same level). The positions of the sources in the secondary astrometric data set are typically known to $\sim 10$ mas. The positions and proper motions are given in a reference frame that is aligned with the International Celestial Reference Frame (ICRF) to better than 0.1 mas at epoch J2015.0, and non-rotating with respect to ICRF to within 0.03 mas yr$^{-1}$.

The photometric data comprises the mean Gaia $G$-band magnitudes for all the sources contained in Gaia DR1, with uncertainties ranging from a few milli-magnitudes at the bright end to $\sim 0.3$ mag at the survey limit (although systematic errors cannot be excluded), as well as light curves for 599 Cepheids and 2595 RR Lyrae variables observed at high cadence around the south ecliptic pole.

We have illustrated the scientific quality of the Gaia DR1 and have also pointed out the substantial shortcomings and the preliminary nature of this first Gaia data release. When using the data presented here the warnings given in Sect. 6 should be considered carefully. However, we are confident of the overall quality of the data, which represents a major advance in terms of available precise positions, parallaxes, proper motions, and homogeneous all-sky photometry. In addition, the scientific exploitation of the data at this early stage will surely improve the quality of future Gaia data releases.

We note in closing that all of the shortcomings listed in this and the accompanying Gaia DR1 papers will be addressed in
future Gaia data releases with very substantial improvements already ready for Gaia DR2.

Acknowledgements. This work has made use of data from the European Space Agency (ESA) space mission Gaia, the data from which were processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement. The Gaia mission website is http://www.cosmos.esa.int/gaia. The authors are current or past members of the ESA Gaia mission team and of the Gaia DPAC. This work has received financial support from: the Algerian Centre de Recherche en Astronomie, Astrophysique et Géophysique de Bouzareah Observatory; the Austrian FWF Hertha Firnberg Programme through grants T359, P20046, and P23773; the BEL-gian federal Science Policy Office (BELSPO) through various PROgramme de Développement d’Expériences scientifiques (PRODEX) grants; the Brazil-France exchange programmes FAPESP-COFECUB and CAPES-COFECUB; the Chinese National Science Foundation through grant NSFC 11573054; the Czech-Republic Ministry of Education, Youth, and Sports through grant L.G. 15010; 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-2009-252526 (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 (GE-NIUS) and through grant 264895 for the Gaia Research for European Astronomy in the Framework of the Gaia-ITN network; the European Research Council (ERC) through grant 320360 and through the European Union’s Horizon 2020 research and innovation programme through grant agreement 670519 (Mixing through grant 320360 and through the European Union’s Horizon 2020 research and innovation programme through grant agreement 670519 (Mixing); the French Centre National d’Études spatiales (CNES); the French Space Agency (CNES); the French government through grants ANR-2011-PIANOS-006, ANR-10-IDEEX-0001-02; the French Academy of Science; the French Centre National de la Recherche Scientifique (CNRS) through action “Défi MASTODONS”; the French Centre National d’Études Spatiales (CNES); the French Ministry of Education, Youth, and Sports through grant L.G. 15010; the Danish Ministry of Education, Youth, and Sports through grant L.G. 15010; the Danish Ministry of Science Foundation through grant NSFC 11573054; the Czech-Republic Ministry of Education, Youth, and Sports through grant L.G. 15010; the Danish Ministry of Science Foundation through grant NSFC 11573054; the Czech-Republic Ministry of Education, Youth, and Sports through grant L.G. 15010; the Danish Ministry of Science Foundation through grant NSFC 11573054; the Czech-Republic Ministry of Education, Youth, and Sports through grant L.G. 15010; the Danish Ministry of Science Foundation through grant NSFC 11573054; the Czech-Republic Ministry of Education, Youth, and Sports through grant L.G. 15010; the Danish Ministry of Science Foundation through grant NSFC 11573054.

References

Cromer, M., & Katz, D. 2011, in EAS PS, 45, 181
ESA 1997, The HIPPARCOS and TYCHO catalogues. Astrometric and photometric star catalogues derived from the ESA HIPPARCOS Space Astrometry Mission, ESA SP, 1200
Henden, A., & Munari, U. 2014, Contributions of the Astronomical Observatory Skalnate Pleso, 43, 519

## Appendix A: List of acronyms

Table A.1. List of acronyms used in this paper.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2MASS</td>
<td>Two-Micron All Sky Survey</td>
</tr>
<tr>
<td>AAVSO</td>
<td>American Association of Variable Star Observers</td>
</tr>
<tr>
<td>APASS</td>
<td>AAVSO Photometric All-Sky Survey</td>
</tr>
<tr>
<td>BP</td>
<td>Blue Photometer</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge-Coupled Device</td>
</tr>
<tr>
<td>CTI</td>
<td>Charge Transfer Inefficiency</td>
</tr>
<tr>
<td>DPAC</td>
<td>Data Processing and Analysis Consortium</td>
</tr>
<tr>
<td>ICRF</td>
<td>International Celestial Reference Frame</td>
</tr>
<tr>
<td>IGSL</td>
<td>Initial <em>Gaia</em> Source List</td>
</tr>
<tr>
<td>OBMT</td>
<td>On Board Mission Timeline</td>
</tr>
<tr>
<td>PSF</td>
<td>Point Spread Function</td>
</tr>
<tr>
<td>RP</td>
<td>Red Photometer</td>
</tr>
<tr>
<td>RVS</td>
<td>Radial Velocity Spectrometer</td>
</tr>
<tr>
<td>TCB</td>
<td>Barycentric Coordinate Time</td>
</tr>
<tr>
<td>TDI</td>
<td>Time-Delayed Integration (CCD)</td>
</tr>
<tr>
<td>TGAS</td>
<td><em>Tycho-Gaia</em> Astrometric Solution</td>
</tr>
</tbody>
</table>
Appendix B: Example Gaia archive queries

Tables B.1–B.3 list the queries in Astronomical Data Query Language form that can be submitted to the Gaia archive in order to retrieve the data necessary to reproduce Figs. 3, 4, 5, 9, and 6. The selection on the standard uncertainty in $G$ ignores the contribution of the $G$-band magnitude zero point error. Including this small (~0.003 mag) error term does not alter the query results, except for the selection for Fig. 5.

Table B.1. Minimal queries that can be submitted to the Gaia archive in the Astronomical Data Query Language to retrieve the data necessary to reproduce the HR diagrams and the magnitude distribution in Figs. 3 and 4.

<table>
<thead>
<tr>
<th>Query to reproduce panel a and b of Fig. 3. This results in a table of 43 546 rows listing the Gaia source identifier, the HIPPARCOS number, the values of $M_G$ based on the HIPPARCOS and Gaia DR1 parallaxes respectively, and the value of $(B − V)$ from the HIPPARCOS catalogue (van Leeuwen 2007).</th>
</tr>
</thead>
</table>
| select gaia.source_id, gaia.hip,  
  gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs_gaia,  
  gaia.phot_g_mean_mag+5*log10(hip.plx)-10 as g_mag_abs_hip,  
  hip.b_v  
  from gaiadr1.tgas_source as gaia  
  inner join public.hipparcos_newreduction as hip  
  on gaia.hip = hip.hip  
  where gaia.parallax/gaia.parallax_error >= 5 and  
  hip.plx/hip.e_plx >= 5 and  
  hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and  
  2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 |

<table>
<thead>
<tr>
<th>Query to reproduce panel c of Fig. 3. This results in a table of 74 771 rows listing the Gaia source identifier, the values of $M_G$ based on the Gaia DR1 parallax, and the value of $(B − V)$ from the HIPPARCOS catalogue (van Leeuwen 2007).</th>
</tr>
</thead>
</table>
| select gaia.source_id, gaia.hip,  
  gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs,  
  hip.b_v  
  from gaiadr1.tgas_source as gaia  
  inner join public.hipparcos_newreduction as hip  
  on gaia.hip = hip.hip  
  where gaia.parallax/gaia.parallax_error >= 5 and  
  hip.plx/hip.e_plx >= 5 and  
  hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and  
  2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 |

<table>
<thead>
<tr>
<th>Query to reproduce Fig. 4. This results in a table of 3174 rows listing the Gaia source identifier, and the values of $M_G$ based on the Gaia DR1 and HIPPARCOS (van Leeuwen 2007) parallaxes, respectively.</th>
</tr>
</thead>
</table>
| select gaia.source_id, gaia.hip,  
  gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs_gaia,  
  gaia.phot_g_mean_mag+5*log10(hip.plx)-10 as g_mag_abs_hip  
  from gaiadr1.tgas_source as gaia  
  inner join public.hipparcos_newreduction as hip  
  on gaia.hip = hip.hip  
  where gaia.parallax/gaia.parallax_error >= 5 and  
  hip.plx/hip.e_plx >= 5 and  
  hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and  
  hip.b_v >= 1.0 and hip.b_v <= 1.1 and  
  2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 |
Table B.2. Minimal queries that can be submitted to the Gaia archive in the Astronomical Data Query Language to retrieve the data necessary to reproduce the HR diagram in Fig. 5 as well as the Pleiades parallax histogram in Fig. 9.

<table>
<thead>
<tr>
<th>Query to reproduce Fig. 5. This results in a table of 1 004 207 rows (1 004 204 when including the zero-point uncertainty on $G$ in the selection criteria) listing the Gaia source identifier and the values of $M_G$ and $(G - K_s)$.</th>
</tr>
</thead>
</table>
| select gaia.source_id,  
  gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs,  
  gaia.phot_g_mean_mag-tmass.ks_m as g_min_ks  
 from gaiadr1.tgas_source as gaia  
 inner join gaiadr1.tmass_best_neighbour as xmatch  
 on gaia.source_id = xmatch.source_id  
 inner join gaiadr1.tmass_original_valid as tmass  
 on tmass.tmass_oid = xmatch.tmass_oid  
 where gaia.parallax/gaia.parallax_error >= 5 and ph_qual = 'AAA' and  
  sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux,2)) <= 0.05 and  
  sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux,2)  
  + power(tmass.ks_msigcom,2)) <= 0.05 |

<table>
<thead>
<tr>
<th>Query to carry out a simplistic selection of Pleiades cluster members and reproduce Fig. 9. This results in a table of 164 rows listing the Gaia source identifier and the Gaia parallax.</th>
</tr>
</thead>
</table>
| select gaia.source_id,  
  gaia.parallax  
 from gaiadr1.tgas_source as gaia  
 where contains(point('ICRS',gaia.ra,gaia.dec),circle('ICRS',56.75,24.12,5)) = 1  
 and sqrt(power(gaia.pmra-20.5,2)+power(gaia.pmdec+45.5,2)) < 6.0 |
Table B.3. Minimal queries that can be submitted to the Gaia archive in the Astronomical Data Query Language to retrieve the data necessary to reproduce the HR diagram in Fig. 6. In this case the query is split into three parts.

---

Query to retrieve stars with HIPPARCOS colour indices. This results in a table of 30009 rows listing the Gaia source identifier, the value of $M_G$ based on the Gaia DR1 parallax, the HIPPARCOS (van Leeuwen 2007) value of $(B-V)$, and the value of $v_\perp$.

```sql
select gaia.source_id,
gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
(gaia.parallax >= 10.0 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
(gaia.parallax >= 10.0 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
(gaia.parallax >= 10.0 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
(gaia.parallax >= 10.0 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2))/gaia.parallax*4.74047 as vperp
```

---

Query to retrieve stars with Tycho-2 colour indices. This results in a table of 8983 rows listing the Gaia source identifier, the value of $M_G$ based on the Gaia DR1 parallax, the Tycho-2 value of $(B-V)$ calculated as $0.85(B-V)_T$, and the value of $v_\perp$.

```sql
select gaia.source_id,
gaiadr1.tgas_source as gaia
inner join public.tycho2 as tycho2
on gaia.tycho2_id = tycho2.id
where gaia.parallax/gaia.parallax_error >= 5 and
sqrt(power(tycho2.e_bt_mag,2) + power(tycho2.e_vt_mag,2)) <= 0.05 and
2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
(gaia.parallax >= 10.0 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
(gaia.parallax >= 10.0 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
(gaia.parallax >= 10.0 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2))/gaia.parallax*4.74047 as vperp
```

---

Query to retrieve stars with APASS colour indices. This results in a table of 2144 rows listing the Gaia source identifier, the value of $M_G$ based on the Gaia DR1 parallax, the APASS value of $(B-V)$, and the value of $v_\perp$.

```sql
select gaia.source_id,
gaiadr1.tgas_source as gaia
inner join public.tycho2 as tycho2
on gaia.tycho2_id = tycho2.id
inner join gaiadr1.urat1_best_neighbour as uratxmatch
on gaia.source_id = uratxmatch.source_id
inner join gaiadr1.urat1_original_valid as urat
on uratxmatch.urat1_oid = urat.urat1_oid
where gaia.parallax/gaia.parallax_error >= 5 and
sqrt(power(tycho2.e_bt_mag,2) + power(tycho2.e_vt_mag,2)) > 0.05 and
sqrt(power(tycho2.e_bt_mag,2) + power(tycho2.e_vt_mag,2)) <= 0.05 and
2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05
(gaia.parallax >= 10.0 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
(gaia.parallax >= 10.0 or
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
sqrt(power(tycho2.e_bt_mag,2) + power(tycho2.e_vt_mag,2))/gaia.parallax*4.74047 as vperp
```

---

The queries above are designed to select Gaia source identifiers, Gaia DR1 parallaxes, HIPPARCOS, Tycho-2, and APASS colour indices, respectively, along with their associated uncertainties and a catalogue of other distances to ensure the validity of the colour indices. The queries are carefully constructed to minimize the number of rows selected, ensuring the retrieval is efficient and the data is accurate. The queries are split into three parts to handle the different colour indices: HIPPARCOS, Tycho-2, and APASS.