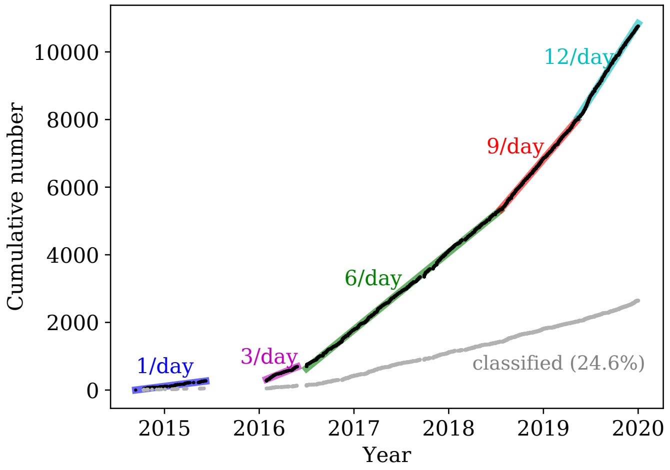

Fig. 10. Cumulative number of alerts detected as a function of publication date. Changes in the alert rate correspond to changes and improvements to AlertPipe, as described in this paper, which allowed us to identify alerts more reliably. Grey line shows the cumulative number of classified alerts, reaching almost 25% of all alerts by the end of 2019 (see Sect. 3.2). Figure includes alerts published from the start of operations until the end of December 2019.

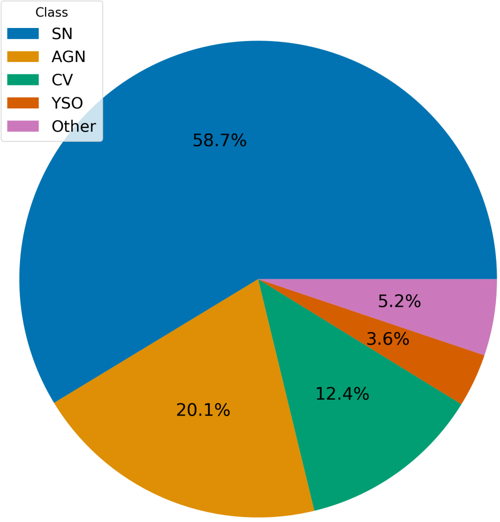

Fig. 11. Distribution of the most common classifications for alerts from GSA. As discussed in the text, there is strong bias in the rate of follow-up and classification in favour of events which look like supernovae or other extragalactic transients.

alerted but were not published at that time (out of a total of 9969 alerts, i.e. 5.6%).

3. Results

3.1. Alert rate

The cumulative number of published alerts is shown in Fig. 10 and shows a number of distinct phases (indicated with different colours in the figure):

- From September 2014 to June 2015: an initial commissioning phase, spanning almost 300 days, where we published alerts at a rate of about one event per day.
- From June 2015 to January 2016: a pause in publication, where we developed most of the filters discussed in Sect. 2.7.1, to minimise the rate of spurious detections.
- January 2016: restarting of operations for AlertPipe and Alert publication. A density map was implemented to ignore events arising in the most crowded regions of sky.
- June 2016: following further improvements to the filters, and removal of the density map restriction.
- April 2018: eyeballing App introduced.
- May 2019: OldSourceSkewVonN detector introduced.

Figure 9 shows the magnitude distribution of alerts detected and published by GSA. Some transients have alerting magnitudes fainter than $G = 19$, our nominal threshold for detection. This can happen for a number of reasons: (1) the first detection of a brightening source is fainter than $G = 19$, but subsequent measurements are brighter, (2) a source which was previously brighter than $G = 19$, fades, (3) in earlier phases of operations we allowed publications of fainter transients.

3.2. Alerts as a function of class

A long-standing problem with transient surveys is the rate of classification compared to the (usually much higher) rate of discovery. For GSA we see that almost 25% of alerts discovered up to the end of 2019 were classified. These classifications were obtained from numerous sources (including Simbad, ADS, TNS, and Astronomical Telegrams as described in Sect. 2.8) and may well be incomplete. The majority of classifications were obtained spectroscopically (and are dominated by SNe). However in the case of microlensing events, a classification could be derived from modelling of the lightcurve alone. An analysis of all events reported to the Transient Name Server in 2019 (Kulkarni 2020), showed that only around 10% of events were classified. This is typically limited by access to ground-based facilities, where the problem scales with magnitude as shown in Fig. 4 of Kulkarni (2020), that is fainter objects are less frequently classified (although they also note that there is a bright tail of events, dominated by GSA, which remain unclassified and are likely stellar in origin). This is supported by Fig. 14, but see also Sect. 4.1.

For the classified GSA alerts, Fig. 11 shows the most common transient classes. We see that supernovae are dominant amongst classified alerts, followed by AGN (this includes QSOs and BL Lac objects) and then CVs. A full list of these broad classifications is given in Table 2, and some illustrative example lightcurves for eight different classified alerts are shown in Fig. 12.

A large fraction of the classifications come from dedicated spectroscopic programmes such as PESSTO (Smartt et al. 2015) and Spectral Energy Distribution Machine (SEDM, Blagorodnova et al. 2018), and therefore they are heavily biased towards supernova discovery by design. Consequently, the class fractions derived from these classifications are not representative of the entire sample of GSA events.

3.3. Spatial distribution

In Fig. 13, we show 9969 *Gaia* alerts published from observations made between 11 June 2016 and 31 December 2019 inclusive. We compare the distribution on the sky (in Galactic coordinates) with three other ongoing transient surveys for the same time range: ASAS-SN (3120 events), Pan-STARRS1 (15 086 events), and ATLAS (7804 events). We note a number of interesting features in the distributions. Firstly, only the *Gaia* and ASAS-SN surveys sample the Galactic plane, while the majority of surveys, including Pan-STARRS1 and ATLAS avoid this

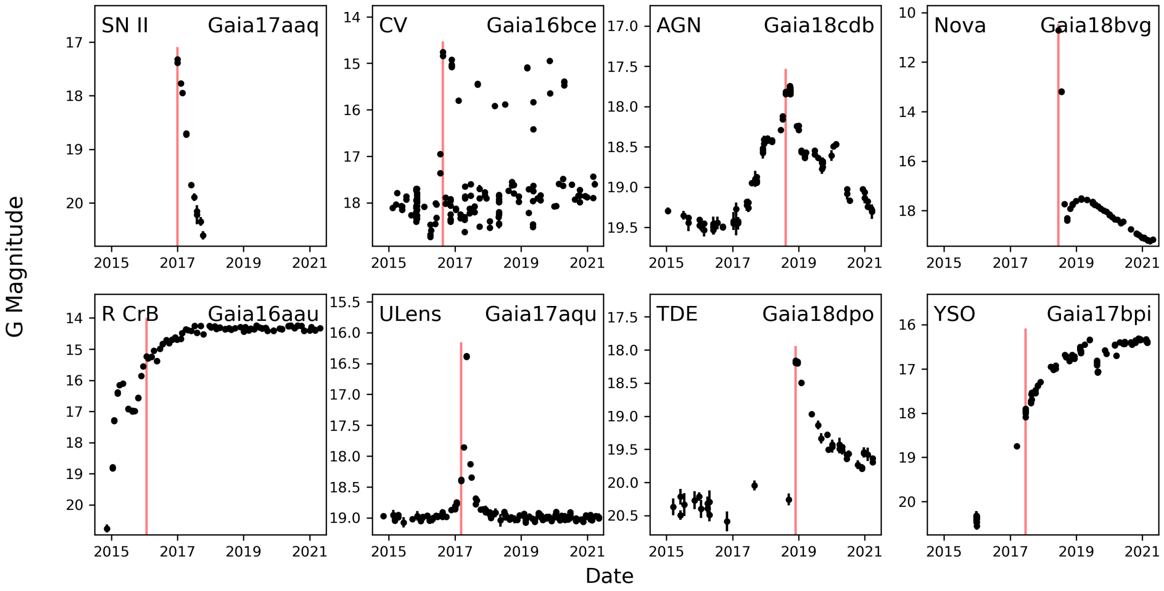

Fig. 12. *Gaia* lightcurves for eight alerts spanning a range of classes. *Top row:* type II supernova, cataclysmic variable, active galactic nucleus, nova. *Bottom row:* R CrB star, microlensing event, tidal disruption event, young stellar object. The red vertical line illustrates the alerting transit for the event. The y-axis gives the *Gaia* magnitude, and the shared x-axis spans mid-2015 to mid-2021.

Table 2. GSA classifications by number and percentage (of classified alerts).

Class	Number	% of Classified
SN	2019	59.0
AGN	717	20.1
CV	442	12.4
YSO	125	3.5
Star	55	1.5
Microlensing	37	1.0
Nova	29	0.8
SLSN	22	0.6
TDE	9	0.2
R CrB	9	0.2
XRB	8	0.2
Symbiotic	7	0.2
SN Impostor	3	0.1
SSO	2	0.1

crowded region. Secondly, only ASAS-SN and GSA are all-sky surveys, the others are based on data taken from a single ground-based observatory, and thus miss a significant fraction of the southern hemisphere. Finally we note that *Gaia* sees an excess of events in the Galactic plane which is not seen by ASAS-SN, perhaps because the latter uses bluer filters (which will be more affected by extinction), is shallower, and has lower spatial resolution.

The overdensity in the plane is further explored in Fig. 14 which shows in two panels the distribution of classified and unclassified *Gaia* alerts. The classified and unclassified alerts are anti-correlated with Galactic latitude. This is perhaps unsurprising, as the main follow-up campaigns (e.g. PESSTO, NUTS) are focused on extragalactic events (such as supernovae and tidal disruption flares), and so avoid the plane by design. We can infer from this that the majority of unclassified alerts are Galactic in origin, and thus the statistics presented in Table 2 are not reflective of the true breakdown of the GSA transient classes.

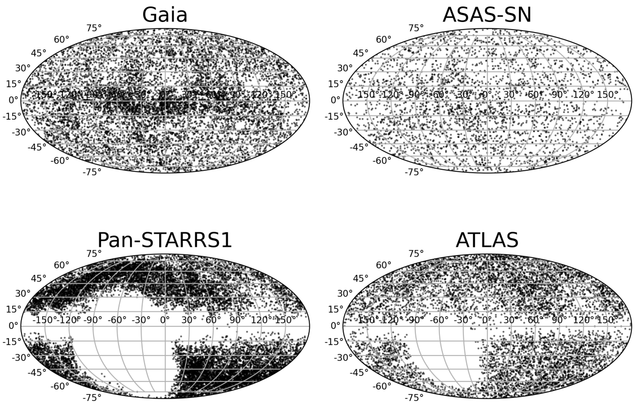

Fig. 13. Survey coverage for GSA (*top-left*), compared to three other transient surveys: ASAS-SN (*top-right*), ATLAS (*bottom-right*) and PanSTARRS (*bottom-left*). Data are presented in Galactic coordinates, with the centre of the Galaxy at the centre of each figure. Transients were all compiled using our local database as described in Sect. 2.8, and reported during the date range June 11 2016 to December 31 2019 inclusive.

3.4. Astrometry

As discussed in Sect. 2 GSA uses astrometry derived by IDT. Previously, Wevers et al. (2019) has shown that the median separation between the *Gaia* alert coordinates and *Gaia* DR2 is 62 milliarcsec (with a standard deviation of 40 milliarcsec). We independently compared the per-transit RA and Dec positions for 10 461 distinct alerts comprising just over 240 000 individual transits, measured between 15 January 2016 and 31 December 2019, with the positions published in *Gaia* DR2. We find the separations between the two coordinate systems are reasonably well described by a Rayleigh Distribution, albeit with a slight excess in the tail to larger separations, presumably arising from systematic differences between the actual and predicted *Gaia* spacecraft attitude (i.e. consistent with spacecraft hits, Fabricius et al. 2016). The best fit model results in an average offset of 55 milliarcsec, with no dependency on magnitude.

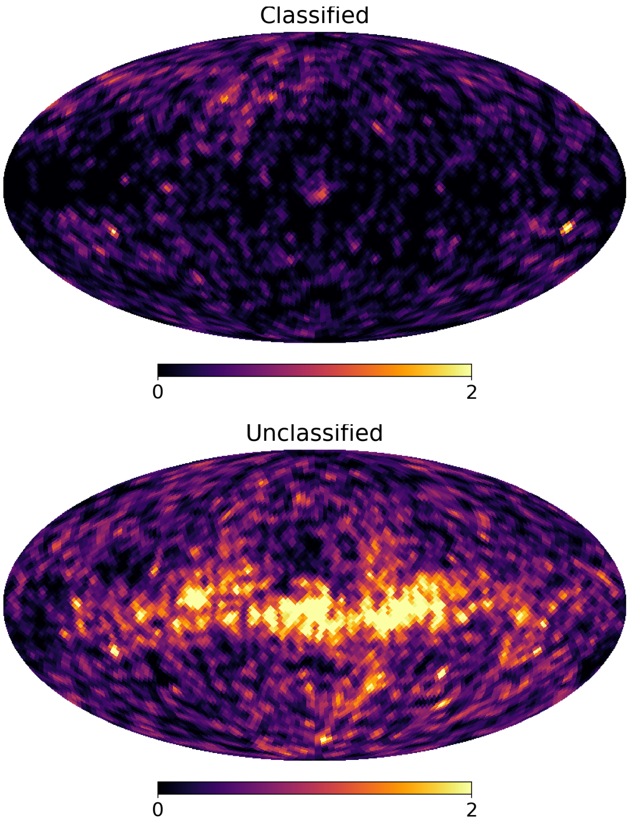

Fig. 14. Distribution on the sky for classified (*upper panel*) and unclassified (*lower panel*) alerts. The map is in Galactic coordinates, using a HEALpix grid with a resolution of about 1.8° (level 5, NSIDE=32, 12 288 pixels), and has been lightly smoothed.

We also note that Yaron et al. (2019) compared the transient positions between published *Gaia* alerts and a number of surveys, finding the following median separations (amongst others): *Gaia*–ZTF: 0.12 arcsec; *Gaia*–Pan-STARRS1: 0.12 arcsec; *Gaia*–ASAS-SN: 1.17 arcsec.

They conclude that GSA astrometric measurements will be treated as ground truth, which is to say that the TNS coordinates will be updated to the *Gaia* positions if and when published.

3.5. Photometry

All transit photometry is calibrated on-the-fly via a database function. The parameters used by the function are derived from the Photometric One Day Calibration (PODC, see Appendix B), and are generated with a one-day cadence. The operation to build the calibration is run roughly monthly, thus at the time of alert, the calibration can be a month out of date (and sometimes more).

To test the precision of PODC, we selected a random sample of 184 000 sources which lie in the SDSS DR7 footprint (avoiding the most crowded regions of the Galactic plane). We required the sources to have a minimum of 10 *Gaia* field-of-view transits, and we used the median of the per-CCD PODC calibrated fluxes as the per-transit CCD flux. We used the standard deviation of multiple transits for a source as a measurement of the precision of a single *Gaia* transit in the Alerts system. In Fig. 15 we show that the precision reaches 1% for sources around $G = 13$, falling to around 10% near the limit of the survey ($G = 20$). Most alerting sources must reach $G = 19$, where the median standard deviation is 0.031 mag.

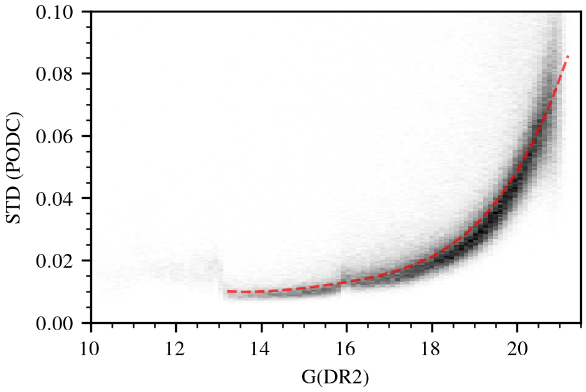

Fig. 15. Standard deviation of the PODC per-transit photometry (in magnitudes), as a function of *Gaia* DR2 magnitude. We overlay a fourth-order polynomial model, fit to the median standard deviation as a function of G_{DR2} magnitude.

We note the features (steps) at $G = 13$ and at $G = 16$, which are consistent with changes to the window-class (see Appendix B for more details). Sources which can swap window class between observations will end up with additional scatter in their photometry, because PODC does not attempt to homogenise the different calibration units. As an approximation, we fitted a model to the median standard deviation versus magnitude:

$$STD = 3.44 - 0.879G + 0.084G^2 - 0.0036G^3 + 0.00006G^4, \quad (2)$$

where G is from DR2, and the model is valid only for the range $13 < G < 21$. For brighter sources, we suggest using a conservative value of 0.02 mag. Discussion on the *accuracy* of the PODC calibration, and comparison with *Gaia* DR2 is included in Appendix B.

3.6. Transient variability in the Hertzsprung–Russell diagram

GSA has a unique feature among other transient surveys: in addition to the G-magnitude, each datapoint (transit) in a *Gaia* transient lightcurve has simultaneous broad-band colour information delivered by the red and blue photometers. Although this colour is essentially uncalibrated for GSA (i.e. derived from the raw pixel samples), it can nevertheless be used to trace the colour evolution of transients as they evolve over time.

Unlike the *Gaia* colours released as part of DR2, at the moment of processing GSA does not have available all the necessary information to accurately calibrate the BP/RP measurements (either in wavelength or in flux). For DR2, this calibration was derived from a large sample of spectrophotometric standard stars within a narrow spectral range, to accurately assess and correct for the relevant distortions (Evans et al. 2018). Here we used uncalibrated colours derived by integrating the raw BP and RP spectra, uncorrected for any throughput variations or wavelength offsets that may have been present. Although this is an approximation, we show that the raw BP–RP colour provides a reasonable assessment of the transient properties.

In order to mitigate the effects of cosmic rays, charge injections and other artefacts that artificially distort the colour, we used a $3\,\sigma$ outlier rejection in the sample values of the spectra before we computed the integrated fluxes. Although this sometimes removed real features in exceptionally strong emission line sources, it significantly improved the overall consistency of the BP–RP colour. For a higher degree of consistency, we performed

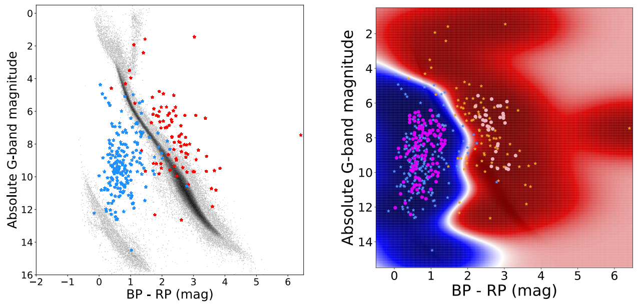

Fig. 16. *Left*: colour-magnitude diagram of *Gaia* DR2 counterparts to *Gaia* Alerts (stars), overlaid on a clean 100 pc sample (grey background) to illustrate the position relative to the main and white dwarf sequences. We colour-code CVs as blue and YSOs as red. *Right*: probability map of a binary SVM classifier (see text). Blue regions indicate CV parameter space, red regions YSOs. The white line is the decision boundary; training samples are shown as coloured stars, while new classifications (with $P > 0.95$) are shown as magenta/pink circles.

a median colour correction such that it agreed with the *Gaia* DR2 colour.

For those alerts that have a parallax measurement in *Gaia* DR2 (which are therefore necessarily restricted to Galactic sources), these observed quantities can be combined to trace their evolution in a colour magnitude diagram (CMD). We used the distance estimates of Bailer-Jones et al. (2018) to calculate the absolute magnitude. Figure 16 shows the sample of CVs and YSOs to illustrate the typical parameter space covered in the CMD. Here, we show the average DR2 colour of each, which already illustrates that while YSOs and CVs might be discriminated by colour to zeroth order, more information (e.g. parallax) is required to provide an accurate separation of the two classes. No correction was made for reddening.

To illustrate the power of colour as well as parallax information, we created a binary classifier using a support vector machine (SVM). We used the standard radial basis function (RBF) kernel in the SCIKIT-LEARN package in Python. Probabilistic output was obtained through 5-fold cross-validation. We used the classified sources as a training set and predicted classifications for 1815 unknown alerts that have a counterpart in DR2. Because we used DR2 colours and absolute magnitudes, the results should be valid more broadly for transients with a DR2 counterpart discovered by other surveys as well. The results are visualised in Fig. 16, where the colour map traces the (binary) classification probability. Blue regions indicate parameter space covered by CVs, while red regions indicate parameter space inhabited by YSOs. The white line indicates the decision boundary between the classes. We overplot a subset of newly classified sources as magenta circles to illustrate the high confidence (probability $P > 0.95$) parameter space for each class. Using this simple algorithm, we classify 638 sources as CVs for $P > 0.95$, while 202 new YSOs are classified. We include a table of these newly classified objects at the CDS[13]. The remainder of the 1815 alerts were not classified with high enough confidence to be included. We caution that this is a very simplistic classifier which uses only the magnitude, colour and parallax of the transients, to show where CVs and YSOs are most likely to be found. Since the colour is available for all our transients, it is a very useful parameter, but the classification will not be perfect. This classifier also only considers two types of objects, so the list may be contaminated with a small number of other objects such as flare stars, variable stars or QSOs. Future *Gaia* data releases, based on more observations, will remove the apparent parallaxes of QSOs included in GDR2.

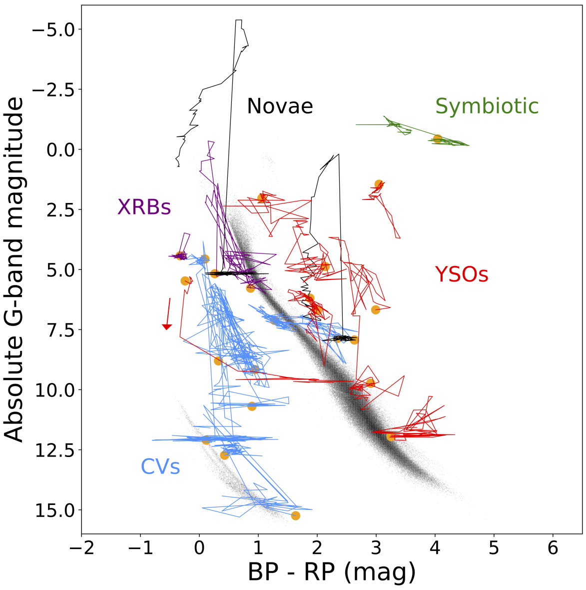

Fig. 17. Colour-magnitude diagram showing the colours per epoch derived from *Gaia* Alerts spectro-photometry, to show the evolution of a selected subsample of sources through the HRD. Orange filled circles mark the beginning of the lightcurve.

Another application of the alert colour information is demonstrated in Fig. 17, where we show a subsample of sources and follow their evolution through the diagram as their properties vary in time. A similar figure is shown in Fig. 11 of Gaia Collaboration (2019), which includes a large sample of periodic and non-periodic variables. We have colour coded sources by classification. Note that some sources show a large spread in BP–RP, indicating that our simple data curation may be insufficient, and a proper spectral calibration is required for more detailed analysis; fully calibrated epoch BP and RP spectra will be released in DR4. Nevertheless, we note that CVs and XRBs have bluer colours in outburst, as expected. We can also see the reddening of emission that occurs in novae between 10 and 100 days post-peak (e.g. Hachisu & Kato 2014), when the wind ejecta expands while the photosphere recedes.

4. Discussion

4.1. Purity

We would like to assess the fraction of the published alerts which are astrophysically real and not due to an artefact or systematic effect in the *Gaia* data. We define this fraction as the purity of our published alerts. While purity is not a quantity frequently reported by surveys, it is important in the context of any population studies based on the survey as well as telescope time spent in the follow up of alerts. As in the rest of this paper we consider those published alerts which were originally detected in or between IDT runs 1046 and 4724 (the last run of 2019, see Sect. 1).

Firstly, we examined how many of our published alerts had been observed by another transient survey, using the Transient Name Server (TNS). This could be either before or after it was

---

[13] The table of classified CVs and YSOs is available at the CDS, and contains the following information. Col. 1: name of the *Gaia* alert, Col. 2: *Gaia* DR2 sourceId, Col. 3: Ra, Col. 4: Dec, Col. 5: parallax, Col. 6: parallax error, Col. 7: *G*-band magnitude, Col. 8: BP–RP colour, Col. 9: classification.

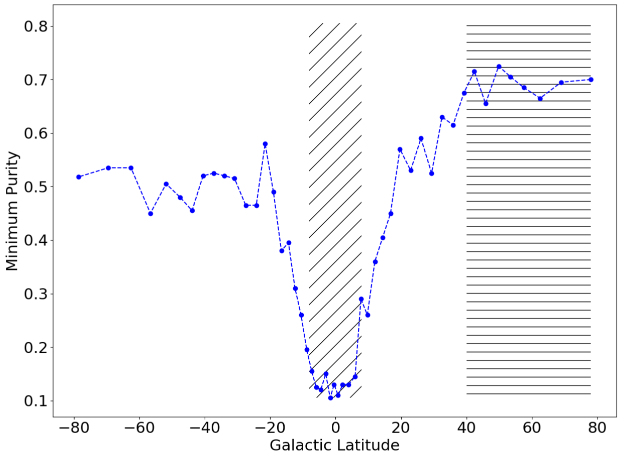

Fig. 18. Minimum purity of the published alert stream as a function of Galactic latitude (see text for details). The behaviour as a function of Galactic latitude may be understood by considering the coverage of other transient surveys which generally avoid the plane, are not full sky and are biased towards northern skies (ASAS-SN is the only other all sky transient survey). Additionally, very few alerts towards the Galactic plane are followed up. The hatched regions show ranges in Galactic latitudes which are used in the purity analysis (Sects. 4.1.1 and 4.1.3). The horizontal hatched region contains 1661 published alerts and has an overall minimum purity of 0.69. The diagonal hatched region covering ±8 degrees around the Galactic plane contains 1628 published alerts and has an overall minimum purity of 0.09.

discovered by *Gaia*. Given that a different survey should not be subject to the same artefacts, we can reliably class these alerts as astrophysically real. Additionally, we consider all alerts which have an ATEL or a spectroscopic classification to be real. The fraction of our published alerts which are real based on this analysis is 0.44. This gives an indication of the minimum level of purity in our sample, as not all alerts were followed up (i.e. had an ATEL or were classified) and most of the other transient surveys are not full sky (with the exception of ASAS-SN).

4.1.1. Purity: dependence on Galactic latitude

Figure 18 shows a strong dependence of the minimum purity level on Galactic latitude. This does not mean however, that our alerts are unreliable in the plane, but rather highlights the absence of coverage by other surveys, as shown in Fig. 13[14], and a low rate of follow-up.

Figure 18 also shows a bias towards northern skies. The purity for positive Galactic latitudes appears higher than for negatives ones, as the majority of ground-based transient survey are based at northern latitudes. A strong dependence on the magnitude of the alert was also found, which is again unsurprising as brighter alerts are more likely to be followed up. It is worth noting that for $b \geq 40°$ and $G_{\rm mag} < 17$ the fraction of our alerts confirmed by ancillary data is 0.93.

4.1.2. Purity: investigation of alert environment

As described in Sect. 2.7 the majority of the causes of spurious alerts were found to be due to environmental effects. Therefore, we chose to study the environment in the vicinity of all the published alerts out to a larger angular separation distance than is possible (due to CPU and timely operation constraints) in our normal processing. The results of our analysis, aimed to detect possible differences between alerts with and without a confirmation in ancillary data, is shown in Fig. 19. Here, we accumulated the relative locations in the AL and AC directions at the time of the alert, of any neighbouring sources in the vicinity out to an angular separation of 2′. In the absence of any environmental effects we should expect a uniform distribution of sources around the location of the alerts. However, this is not what is seen in Fig. 19 where we see an excess in the number of sources at narrow angular separations from the alert. While also visible for the unconfirmed alerts, this excess is more pronounced for the confirmed alerts meaning this environmental effect must be due to the nature of the alerts themselves rather than a spurious detection caused by an environmental effect.

4.1.3. Purity: NewSource versus OldSource alerts

In an effort to understand this, these plots were regenerated, but this time the alerts were divided between NewSource and OldSource alerts. These are shown in Fig. 20, where we see that this excess is barely visible for OldSource alerts while being clearly visible for NewSource alerts. In addition, in Fig. 20, we plot histograms of the number of sources as a function of angular separation out to 20″ to more easily see the location of the excess. As the majority of the NewSource alerts are SNe, the reason for the excess becomes clear; these are sources associated with the host galaxy. The angular extent of many galaxies is such that the on-board detection may record multiple detections at various regions of brighter emission in the galaxy. It is these sources which are being found by the environmental search, and that result in the excess at narrow angular separations. As evidenced by Fig. 20, OldSource alerts are not completely immune to this effect as SNe are occasionally detected via the OldSource route as a brightening of a previously observed source associated with a galaxy. This also explains the difference in the size of this effect between the confirmed and unconfirmed alerts in Fig. 19 as more SNe are followed-up and hence classified than other classes of transients leading to the more pronounced excess for the confirmed alerts.

Importantly, this demonstrates the effectiveness of this technique, of searching in the vicinity of all the published alerts and accumulating their neighbouring sources as a function of their AL and AC positions at the time of the alert. It also demonstrates that there are no other visible environment effects. As an additional check we reproduced the plots of accumulated sources in the AL and AC directions for the alerts in the horizontal hatched and diagonal hatched regions in Fig. 18, corresponding to high ($b \geq 40°$) and low ($|b| < 8°$) Galactic latitude regions. We find a completely uniform distribution of sources at low Galactic latitudes and that the excess is caused by the host galaxies at high Galactic latitudes. Finally, we split the alerts based on their magnitude at the time of detection, and see essentially the same behaviour for the bright and faint alerts.

4.1.4. Purity: summary

It is now possible to conclude that we find absolutely no evidence for any remaining environmentally induced spurious alerts in our sample of published alerts. There is no evidence that the published alerts in the Galactic plane are any less reliable than those at higher Galactic latitudes, nor that fainter alerts are any less reliable than brighter ones. We conclude that the overall purity of our published alerts is comparable to the subset where

---

[14] To see the coverage footprints of the other surveys in TNS go to `https://www.wis-tns.org/stats-maps/maps`

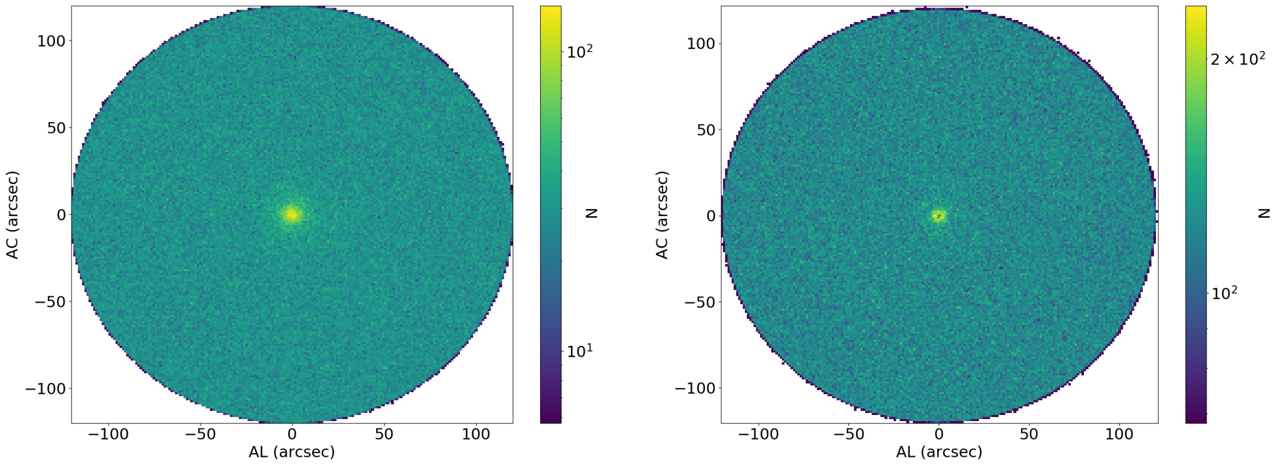

Fig. 19. Number of neighbouring sources within 2′ of each published alert in the AL and AC directions at the time of the alert, accumulated for all alerts which are confirmed by ancillary data (*left*), and for the remaining alerts (*right*). Note that all sources within 0.2″ of the alert positions are excluded as they are deemed to belong to the alerts themselves. The distribution of sources around the alerts shows no excess in the AL and AC directions, as might be expected were there some residual contamination due to the PSF spikes of bright stars causing false detections. There is an excess in the number of sources at very close angular separations, but this is more pronounced for the confirmed alerts than those which are not confirmed by ancillary data.

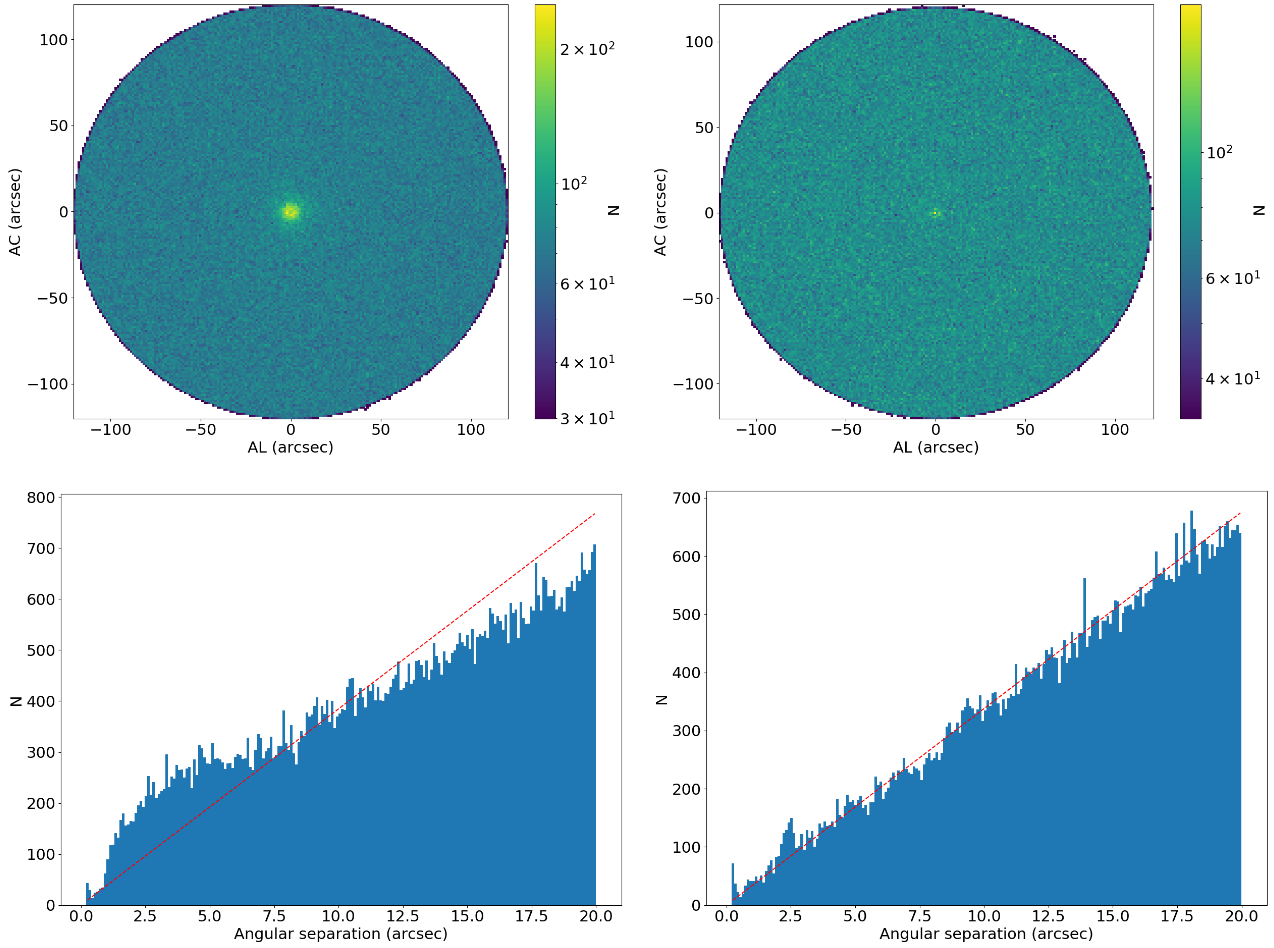

Fig. 20. *Top*: number of neighbouring sources within 2′ of each published alert in the AL and AC directions at the time of the alert, accumulated for all NewSource alerts (*left*), and for all OldSource alerts (*right*). Again all sources within 0.2″ of the alert positions are excluded as they are deemed to belong to the alerts themselves. *Bottom*: number of neighbouring sources as a function of angular separation out to 20″ around the published alert, accumulated for all NewSource alerts (*left*), and for all OldSource alerts (*right*). The red dashed line is the expectation for the number of sources based on the assumption of the same number of sources per unit area. Here we can see that the excess in the number of sources at very close angular separations, is much more pronounced for NewSource alerts than for OldSource alerts. The range of angular separations in which there is an excess in the number of neighbouring source is more clearly seen in the bottom panels. A small excess is still visible around 2.5″ for OldSource alerts.

$b \geq 40°$ and $G_{\rm mag} < 17$, and the fraction of our alerts confirmed by ancillary data is 0.93.

4.2. Completeness

Completeness of a transient survey is a simple concept, but is hard to measure in practice. To be able to measure event rates, it is important to understand what fraction of transients we miss. We have demonstrated that GSA is a transient survey with high purity, however we may expect this to come at the price of completeness.

To quantify this, we have considered a sample of reasonably long lived, bright and well understood transients which *Gaia* has had a good chance of detecting. Of course, we have imperfect knowledge of what the Universe really looks like, and comparisons with external transient surveys can only be as complete as the reference material. We also comment that this approach says little about *Gaia*'s completeness to short-lived transient events such as outbursts from Cataclysmic Variables, or flares from M dwarfs. These more complex selection functions are not considered in this paper.

For our completeness study we use TNS as a starting point, which also records spectroscopic classifications for significant numbers of SNe. From 11 July 2016 until 31 December 2019, a period through which GSA has been operating in a stable mode, there were a total of 5367 classified SNe reported to TNS across all magnitudes. We restrict ourselves to a subset of 2826 SNe to which GSA should be sensitive, with a reported magnitude of $m = 19$ or brighter, noting that these come from a large variety of independent transient surveys, with observations made in different photometric systems.

From this set, we note that 1314 were observed and reported by GSA (379 were reported first by GSA, 39%). A small number of GSA detections (six) came through our salvaging process (see Sect. 2.7.7), the rest were entirely independent detections. It may be that GSA would have detected them following another scan, however in this analysis we consider them to be non-detections by *Gaia*, to be as conservative as possible. Our overall external completeness $C_{\rm E} = 1308/2826 = 0.46$.

4.2.1. Completeness: scanning law

We expect that the largest contributor to the GSA missing events will be the requirement to have two detections from distinct fields-of-view, and separated by less than 40 days. Figure 21 shows the distribution in the number of scans by *Gaia* of the selected 2826 SNe within 40 days of the event notification date, for both the events detected and missed by GSA. This figure supports our expectations, and shows that the internal completeness $C_{\rm I} = 0.57$ where we have exactly two scans. The median completeness for N(scans) >2 is 0.8, implying that the *Gaia* scanning law, and the need to minimise the false alarm rate, dominates the completeness of GSA. For all subsequent analysis, we only include SNe for which there are at least two *Gaia* scans within 40 days of the trigger. A total of 1073 SNe were detected in this subset, and 491 were not, thus our overall internal completeness $C_{\rm I} = 0.69$ on average.

4.2.2. Completeness: magnitude

Figure 22 shows the dependence of GSA detection on SN magnitude. The apparent incompleteness for bright SNe is dominated by small number statistics. Moving towards fainter magnitudes, there does seem to be a roll-off in the completeness which falls from 0.75 (±0.08, G = 16–17) at peak, to 0.60 (±0.03, G = 18–19). This may be in part due to the distinct filters used by the surveys, but also because the SNe, in some cases, may already be declining at the time of announcement.

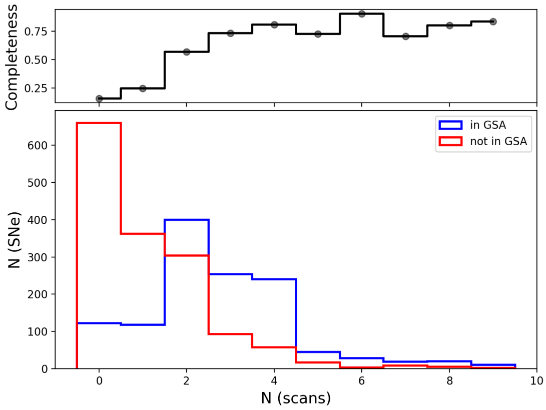

Fig. 21. *Main panel:* histograms of the numbers of SNe reported to TNS between 11 July 2016 and 31 December 2019, as a function of the number of times *Gaia* scanned the location of the event from the date of detection of the event until 40 days after. The histograms are divided into two samples: blue independently detected and published by GSA, red not detected and published by GSA. *Upper panel:* fraction of the total number detected by GSA. Note that some of the SNe were first detected by *Gaia* more than 40 days after the event was reported to TNS, thus the N(scans) = 0, 1 bins are occupied.

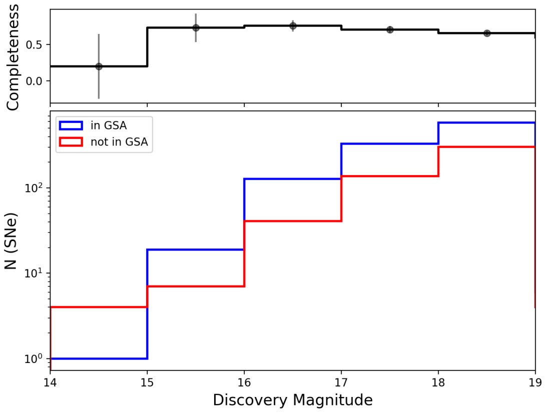

Fig. 22. Completeness as a function of magnitude for SNe with 2 or more *Gaia* scans. *Main panel*: histograms are divided into two samples: blue independently detected and published by GSA, red not detected and published by GSA. *Upper panel*: fraction of the total number detected by GSA.

4.2.3. Completeness: nuclear events

Finally, we explored completeness versus separation from the host galaxy, and this is shown in Fig. 23. The sample of 'host' galaxies was created by cross-matching the coordinates of the SN sample (again only those with ≥2 scans are considered) against the 2MASS extended source list. There is a cluster of datapoints at separations larger than ∼30″ at the bottom-right of Fig. 23. These are possibly mismatches and suggestive of incompleteness in the galaxy sample. There is also a very clear deficit of GSA SN discoveries within 3″ of the centre of galaxies (between 3″ and 80″ the average completeness is 0.79). A similar deficit has previously been ascribed (at least in part) to source

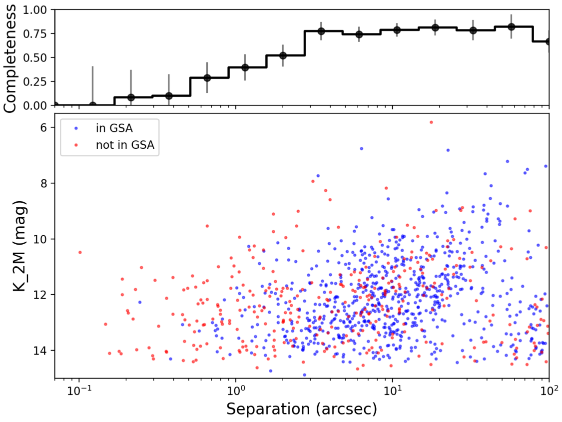

Fig. 23. *Main panel*: 2MASS 'host' galaxy magnitude (*K*-band) versus distance to SN (arcseconds) for those events published by GSA (in blue), and those missed (in red). *Upper panel*: completeness as a function of separation. We also note the likely incorrect host identification for separations greater than 30″.

confusion in the cross-match phase of IDT for transients in the centres of galaxies. Kostrzewa-Rutkowska et al. (2018) found that around 45% of nuclear transients were missed by GSA for this reason. Our completeness for SNe appears to be even lower than this in the nuclear region (though note the significant error bars). Our result also contrasts with simulations (Blagorodnova et al. 2016), which predicted that 90% of SNe would be resolved from their host galaxies by *Gaia*, and detected as NewSources for separations larger than 0.3″, provided that the magnitude of the SN is comparable to the galaxy's bulge.

It is also worth comparing these results to the discussion on purity in Sect. 4.1. Particularly in the bottom left-hand panel of Fig. 20 we can see a significant excess of NewSources associated with neighbours. This is presumably dominated by SNe associated with galaxies, although there is likely a contribution from Cataclysmic Variables in crowded environments (the Galactic plane). This excess falls sharply at close separations, around 1″, indicative of a reduction in the sensitivity of the NewSource detector to complex environments.

5. Summary

In this paper, we have described the *Gaia* Photometric Science Alerts system covering the first 5 yr of operations. Our ingestion and processing system handles upwards of 60 million transits per day, searching for new sources, and sources which change significantly in brightness (both upwards and downwards). The flux-change detectors work with the lightcurves, and thus are sensitive not only to sudden changes in brightness, but also to slowly varying sources.

A series of (mostly environmental) filters is applied to reduce the detection rate from a few thousand provisional candidates to a shortlist of several tens of events per day. These checks also identify likely long-period variables (such as Miras) and known Solar System objects. Surviving candidates are subjected to human scrutiny before being published via our Alerts Website, TNS discovery reports and VOEvents.

When an alert is published, all available internally calibrated *Gaia G* band photometry, and uncalibrated BP/RP spectra, of that object becomes public. This includes subsequent measurements of the same object, with the data added to the alert webpages when processed.

We currently publish alerts at a rate of approximately 12 events day^{-1}, and almost 25% of them are ultimately classified. The published classifications are dominated by SNe, but we show that this is biased by the extensive supernova follow-up campaigns. The bulk of our unclassified alerts reside in the Galactic plane, and are therefore likely to be Galactic in origin. As an experiment, we built a simple probabilistic alert classifier using uncalibrated BP–RP colour and *Gaia* DR2 parallax (where available), leading to the identification of 638 new candidate CVs and 202 new candidates YSOs. We also show that per-transit data (*G*-band and uncalibrated BP and RP colours) for alerting sources with available DR2 parallaxes can be used to trace the evolution of the transient's position in the colour-magnitude diagram, revealing its nature without the need for spectroscopic confirmation.

We investigated the astrometry of GSA and showed that the accuracy of individual alert detections is 55 milliarcsec when compared to *Gaia* DR2, and is independent of source magnitude. The photometry of our alerts has a precision of 1% for sources around $G = 13$, falling to around 10% at $G = 20.7$. Most alerting sources must reach $G = 19$, where the median standard deviation is 0.031 mag.

GSA suffers very low levels of contamination from false positives. We showed that the minimum purity of the survey in uncrowded regions for sources with $G < 17$ is 93%. Ancillary data is harder to find in the Galactic plane, and for fainter sources, but we find no evidence of additional sources of contamination at faint magnitudes, or in crowded regions.

We also investigated the completeness of the survey, by comparing *Gaia* discoveries to classified supernovae published via TNS (across all sub-types). We measure an overall completeness of 46%, and find that the *Gaia* scanning law, combined with our requirement for two *Gaia* detections, dominates our ability to detect a supernova. Our completeness for supernovae with two or more scans is 79%, unless within 3 arcsec of the nucleus of the host galaxy, where it drops significantly.

Finally, we note that a total of 2612 alerts spanning observations taken between 25 July 2014 and 28 May 2017 will be included in *Gaia* DR3 in a supplementary table.

Acknowledgements. We thank the anonymous referee for comments and suggestions that improved this article. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. Further details of funding authorities and individuals contributing to the success of the mission is shown at `https://gea.esac.esa.int/archive/documentation/GEDR3/Miscellaneous/sec_acknowl/`. We thank the United Kingdom Particle Physics and Astronomy Research Council (PPARC), the United Kingdom Science and Technology Facilities Council (STFC), and the United Kingdom Space Agency (UKSA) through the following grants to the University of Bristol, the University of Cambridge, the University of Edinburgh, the University of Leicester, the Mullard Space Sciences Laboratory of University College London, and the United Kingdom Rutherford Appleton Laboratory (RAL): PP/D006511/1, PP/D006546/1, PP/D006570/1, ST/I000852/1, ST/J005045/1, ST/K00056X/1, ST/K000209/1, ST/K000756/1, ST/L006561/1, ST/N000595/1, ST/S000623/1, ST/N000641/1, ST/N000978/1, ST/N001117/1, ST/S000089/1, ST/S000976/1, ST/S001123/1, ST/S001948/1, ST/S002103/1, and ST/V000969/1. This paper made use of the Whole Sky Database (WSDB) created by Sergey Koposov and maintained at the Institute of Astronomy, Cambridge with financial support from the Science and Technology Facilities Council (STFC) and the European Research Council (ERC). We thank the *William Herschel* and *Isaac Newton* Telescopes on the Roque de los Muchachos Observatory, La Palma, Spain, as well as the Optical Infrared Coordination Network for Astronomy (OPTICON) for their support of this project through telescope time, especially during the commissioning and verification phases. We thank the *Copernico* 1.82 m telescope (Mt. Ekar, Asiago Italy) operated by INAF Padova for supporting the project through telescope time (under the Large Programme Tomasella-SNe) during the verification phases. We acknowledge observations taken as part of the PESSTO project collected at the European Organisation for Astronomical Research in the Southern Hemisphere under ESO programme 199.D-0143. Authors at the ICCUB were supported by the Spanish Ministry of Science, Innovation and University (MICIU/FEDER, UE) through grant RTI2018-095076-B-C21, and the Institute of Cosmos Sciences University of Barcelona (ICCUB, Unidad de Excelencia 'María de Maeztu') through grant CEX2019-000918-M. This work is supported by Polish NCN grants: Daina No. 2017/27/L/ST9/03221, Harmonia No. 2018/30/M/ST9/00311, Preludium No. 2017/25/N/ST9/01253 and MNiSW grant DIR/WK/2018/12 as well as the European Commission's Horizon2020 OPTICON grant No. 730890. The Authors would like to thank the Warsaw University OGLE project for their continuous support in this work. AB acknowledges financial support from the Netherlands Research School for Astronomy (NOVA). AG acknowledges the financial support from the Slovenian Research Agency (grants P1-0031, I0-0033, J1-8136, J1-2460). AH was funded in part by the Leverhulme Trust through grant RPG-2012-541 and by the European Research Council grant 320360. AP acknowledges support from the NCN grant no. 2016/21/B/ST9/01126. CM acknowledges support from Jim and Hiroko Sherwin. DAK acknowledges support from the Spanish research projects AYA 2014-58381-P, AYA2017-89384-P, from Juan de la Cierva Incorporación fellowship IJCI-2015-26153, and from Spanish National Research Project RTI2018-098104-J-I00 (GRBPhot). EB and STH are funded by the Science and Technology Facilities Council grant ST/S000623/1. TW was funded in part by European Research Council grant 320360 and by European Commission grant 730980. GC acknowledges the Agenzia Spaziale Italiana (ASI) for its continuing support through contract 2018-24-HH.0 to the Italian Istituto Nazionale di Astrofisica (INAF). GD acknowledges the observing grant support from the Institute of Astronomy and Rozhen NAO BAS through the bilateral joint research project "*Gaia* Celestial Reference Frame (CRF) and fast variable astronomical objects" (during 2020-2022, leader is G. Damljanovic), and support by the Ministry of Education, Science and Technological Development of the Republic of Serbia (contract no. 451-03-68/2020-14/200002). G. Marton acknowledges support from the EC Horizon 2020 project OPTICON (730890) and the ESA PRODEX contract no. 4000129910. MF is supported by a Royal Society - Science Foundation Ireland University Research Fellowship NB acknowledges support from the research programme VENI, with project number 016.192.277, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). NI is partially supported by Polish NCN DAINA grant no. 2017/27/L/ST9/03221. PAW acknowledges research funding from the South African National Research Foundation. RWW was funded by the Science and Technology Facilities Council grant ST/P000541/1. V.A.R.M.R. acknowledges financial support from Radboud Excellence Initiative, the Fundação para a Ciência e a Tecnologia (FCT) in the form of an exploratory project of reference IF/00498/2015/CP1302/CT0001, FCT and the Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020, and supported by Enabling Green E-science for the Square Kilometre Array Research Infrastructure (ENGAGE-SKA), POCI-01-0145-FEDER-022217, and PHOBOS, POCI-01-0145-FEDER-029932, funded by Programa Operacional Competitividade e Internacionalização (COMPETE 2020) and FCT, Portugal. ZKR acknowledges funding from the Netherlands Research School for Astronomy (NOVA). ZN acknowledges support from the ESA PRODEX contract nr. 4000129910.

References

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, AJ, 156, 58
Baltay, C., Rabinowitz, D., Hadjiyska, E., et al. 2013, PASP, 125, 683
Bastian, U. 2013, GAIA-C3-TN-ARI-BAS-020
Bellm, E. C. 2016, PASP, 128, 1
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002
Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, MNRAS, 478, 611
Berthier, J., Vachier, F., Thuillot, W., et al. 2006, Astronomical Data Analysis Software and Systems XV, eds. C., Gabriel, C., Arviset, D., Ponz, & S., Enrique, 351, 367
Blagorodnova, N., Koposov, S. E., Wyrzykowski, Ł., Irwin, M., & Walton, N. A. 2014, MNRAS, 442, 327
Blagorodnova, N., Van Velzen, S., Harrison, D. L., et al. 2016, MNRAS, 455, 603
Blagorodnova, N., Neill, J. D., Walters, R., et al. 2018, PASP, 130, 035003
Boch, T., & Fernique, P. 2014, in Astronomical Society of the Pacific Conference Series, 485, Astronomical Data Analysis Software and Systems XXIII, eds. N. Manset, & P. Forshay, 277
Bose, S., Dong, S., Pastorello, A., et al. 2018, ApJ, 853, 57
Boubert, D., Guillochon, J., Hawkins, K., et al. 2018, MNRAS, 479, 2789
Boubert, D., Everall, A., & Holl, B. 2020, MNRAS, 497, 1826
Campbell, H. C., Marsh, T. R., Fraser, M., et al. 2015, MNRAS, 452, 1060
Cannizzaro, G., Fraser, M., Jonker, P. G., et al. 2020, MNRAS, 493, 477
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints, [arXiv:1612.05560]
Cieza, L. A., Ruíz-Rodríguez, D., Perez, S., et al. 2018, MNRAS, 474, 4347
de Bruijne, J. H. J., Allen, M., Azaz, S., et al. 2015, A&A, 576, A74
Delgado, A., Hodgkin, S., Evans, D. W., et al. 2019a, in Astronomical Society of the Pacific Conference Series, 523, Astronomical Data Analysis Software and Systems XXVII, eds. P. J. Teuben, M. W. Pound, B. A. Thomas, & E. M. Warner, 261
Delgado, A., Rixon, G., van Leeuwen, G., et al. 2019b, in Astronomical Society of the Pacific Conference Series, 521, Astronomical Data Analysis Software and Systems XXVI, eds. M. Molinaro, K. Shortridge, & F. Pasian, 507
Dong, S., Bose, S., Chen, P., et al. 2017, ATel, 10498, 1
Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, ApJ, 696, 870
Evans, D. W., Riello, M., De Angeli, F., et al. 2018, A&A, 616, A4
Fabricius, C., Bastian, U., Portell, J., et al. 2016, A&A, 595, A3
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration (Eyer, L., et al.) 2019, A&A, 623, A110
Gentile Fusillo, N. P., Tremblay, P.-E., Gänsicke, B. T., et al. 2019, MNRAS, 482, 4570
Gomez, S., Berger, E., Nicholl, M., et al. 2019, ApJ, 881, 87
Green, M. J., Marsh, T. R., Steeghs, D. T. H., et al. 2018, MNRAS, 476, 1663
Hachisu, I., & Kato, M. 2014, ApJ, 785, 97
Hillenbrand, L. A., Contreras Peña, C., Morrell, S., et al. 2018, ApJ, 869, 146
Hillenbrand, L. A., Reipurth, B., Connelley, M., Cutri, R. M., & Isaacson, H. 2019, AJ, 158, 240
Hodapp, K. W., Reipurth, B., Pettersson, B., et al. 2019, AJ, 158, 241
Holl, B., Audard, M., Nienartowicz, K., et al. 2018, A&A, 618, A30
Holoien, T. W. S., Prieto, J. L., Stanek, K. Z., et al. 2014, ApJ, 785, L35
Kaiser, N., Burgett, W., Chambers, K., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 7733, Ground-based and Airborne Telescopes III, eds. L. M. Stepp, R. Gilmozzi, & H. J. Hall, 77330E
Kangas, T., Elias-Rosa, N., Lundqvist, P., et al. 2016, ATel, 9071, 1
Kashi, A., Michaelis, A. M., & Feigin, L. 2019, Galaxies, 8, 2
Koposov, S., & Bartunov, O. 2006, in Astronomical Society of the Pacific Conference Series, 351, Astronomical Data Analysis Software and Systems XV, eds. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 735
Kostrzewa-Rutkowska, Z., Jonker, P. G., Hodgkin, S. T., et al. 2018, MNRAS, 481, 307
Kostrzewa-Rutkowska, Z., Jonker, P. G., Hodgkin, S. T., et al. 2020, MNRAS, 493, 3264
Kozłowski, S., Udalski, A., Wyrzykowski, Ł., et al. 2013, Acta Astron., 63, 1
Kulkarni, S. R. 2020, PASP, submitted [arXiv:2004.03511]
Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, PASP, 121, 1395
Lemon, C. A., Auger, M. W., & McMahon, R. G. 2019, MNRAS, 483, 4242
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
Lipunov, V., Kornilov, V., Gorbovskoy, E., et al. 2010, Adv. Astron., 2010, 349171
Makarov, D., Prugniel, P., Terekhova, N., Courtois, H., & Vauglin, I. 2014, A&A, 570, A13
Marton, G., Ábrahám, P., Szegedi-Elek, E., et al. 2019, MNRAS, 487, 2522
Mignard, F. 2016, Gaia Technical Note GAIA-C4-TN-OCA-FM-056-1
Mignard, F., Bailer-Jones, C., Bastian, U., et al. 2008, in A Giant Step: from Milli- to Micro-arcsecond Astrometry, ed. W. J. Jin, I. Platais, & M. A. C. Perryman, 248, 224–230
Nicholl, M., Berger, E., Margutti, R., et al. 2017, ApJ, 835, L8
Paczynski, B. 1996, ARA&A, 34, 419
Riello, M., De Angeli, F., Evans, D. W., et al. 2018, A&A, 616, A3
Rixon, G., Fraser, M., Koposov, S., et al. 2014, ATel, 6593, 1
Roelens, M., Eyer, L., Mowlavi, N., et al. 2017, MNRAS, 472, 3230
Roelens, M., Eyer, L., Mowlavi, N., et al. 2018, A&A, 620, A197
Seabroke, G., Cropper, M., Baker, S., et al. 2020, ArXiv e-prints, [arXiv:2010.16337]
Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48
Sicilia-Aguilar, A., Oprandi, A., Froebrich, D., et al. 2017, A&A, 607, A127
Smart, R. L., & Nicastro, L. 2014, A&A, 570, A87
Smartt, S. J., Valenti, S., Fraser, M., et al. 2015, A&A, 579, A40
Staley, T. D., & Fender, R. 2016, ArXiv e-prints, [arXiv:1606.03735]
Szegedi-Elek, E., Ábrahám, P., Wyrzykowski, Ł., et al. 2020, ApJ, 899, 130
Torra, F., Castañeda, J., Fabricius, C., et al. 2021, A&A 649, A10 (*Gaia* EDR3 SI)
von Neumann, J. 1941, Ann. Math. Statist., 12, 367
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Wevers, T., Jonker, P. G., Hodgkin, S. T., et al. 2018, MNRAS, 473, 3854

Wevers, T., Pasham, D. R., van Velzen, S., et al. 2019, MNRAS, 488, 4816
Wyrzykowski, Ł., Kostrzewa-Rutkowska, Z., Kozłowski, S., et al. 2014, Acta Astron., 64, 197
Wyrzykowski, Ł., Mróz, P., Rybicki, K. A., et al. 2020, A&A, 633, A98
Xiang, D., Song, H., Wang, X., et al. 2017, ATel, 10442, 1
Yaron, O., Gal-Yam, A., Ofek, E., Sass, A., & Knezevic, N. 2019, Transient Name Server AstroNote, 15, 1
Young, K. E., Young, C. H., Lai, S.-P., Dunham, M. M., & Evans, Neal J., I. 2015, AJ, 150, 40

---

[1] Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK.
e-mail: `sth@ast.cam.ac.uk`; `dlh@ast.cam.ac.uk`
[2] Kavli Institute for Cosmology, Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
[3] European Southern Observatory, Alonso de Córdova 3107, Vitacura, Casilla 19001, Santiago, Chile
[4] RHEA Group for ESA/ESAC, Camino bajo del Castillo s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[5] Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands
[6] SRON, Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands
[7] Warsaw University, Astronomical Observatory, Department of Physics, Al. Ujazdowskie 4, 00-478, Warszawa, Poland
[8] Department of Astrophysics/IMAPP, Radboud University, PO Box 9010, 6500 GL Nijmegen, The Netherlands
[9] Department of Physics, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK
[10] School of Physics, O'Brien Centre for Science North, University College Dublin, Belfield, Dublin 4, Ireland
[11] Isaac Newton Group of Telescopes, Apdo. 321, E-38700 Santa Cruz de la Palma, Canary Islands, Spain
[12] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[13] Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Eötvös Loránd Research Network, Konkoly Thege 15-17, 1121 Budapest, Hungary
[14] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, France
[15] Departament de Física Quàntica i Astrofísica, Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí Franquès 1, 08028 Barcelona, Spain
[16] European Space Agency (ESA), European Space Research and Technology Centre (ESTEC), Keplerlaan 1, 2201AZ, Noordwijk, The Netherlands
[17] STFC, Rutherford Appleton Laboratory, Harwell, Didcot OX11 0QX, UK
[18] Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
[19] ELTE Eötvös Loránd University, Institute of Physics, Pázmány Péter sétány 1/A, 1117 Budapest, Hungary
[20] INAF – Osservatorio Astronomico di Roma, Via Frascati 33, 00078, Monte Porzio Catone (Roma), Italy
[21] Space Science Data Center – ASI, Via del Politecnico SNC, 00133 Roma, Italy
[22] ARI/ZAH, Heidelberg University, Heidlelberg, Germany
[23] School of Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK
[24] Centre for Advanced Instrumentation, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK
[25] DAPCOM for Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
[26] School of Physical Sciences, The Open University, Walton Hall, Milton Keynes MK7 67AA, UK
[27] INAF - Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Piero Gobetti 93/3, 40129 Bologna, Italy
[28] Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK
[29] Astronomical Observatory (street Volgina 7, 11060 Belgrade, Serbia)
[30] National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314, USA
[31] Sorbonne Universités, UPMC Univ Paris 6 et CNRS, UMR 7095, Institut d'Astrophysique de Paris, IAP, 75014 Paris, France
[32] Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK
[33] Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain
[34] SUPA, School of Physics & Astronomy, University of St Andrews, North Haugh, St Andrews, KY16 9SS, UK
[35] Department of Astronomy, University of Geneva, Chemin Pegasi 51, 1290 Versoix, Switzerland
[36] Center for Astrophysics and Cosmology, University of Nova Gorica, Vipavska 13, 5000 Nova Gorica, Slovenia
[37] University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada
[38] Centre for Astrophysics and Planetary Science, School of Physical Sciences, University of Kent, Canterbury CT2 7NH, UK
[39] ESA, ESAC, Camino Bajo del Castillo s/n, Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Spain
[40] Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[41] Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, 18008 Granada, Spain
[42] School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK
[43] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53111 Bonn, Germany
[44] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[45] DiRAC Institute, Department of Astronomy, University of Washington, Seattle, WA 98195, USA
[46] INAF-Osservatorio Astrofisico di Catania, via S. Sofia 78, 95123 Catania, Italy
[47] Faulkes Telescope Project, School of Physics and Astronomy, Cardiff University, The Parade, Cardiff, CF24 3AA, Wales, UK
[48] Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA
[49] Center for Data Driven Discovery, California Institute of Technology, Pasadena, CA 91125, USA
[50] Department of Physics, University of Bath, Claverton Down, Bath, BA2 7AY, UK
[51] ZARM, Center of Applied Space Technology and Microgravity, Bremen University, Germany
[52] Ruđer Bošković Institute, Bijenička cesta 54, Zagreb, Croatia
[53] Instytut Astronomiczny, Uniwersytet Wrocławski, Kopernika 11, 51-622 Wrocław, Poland
[54] South African Astronomical Observatory, PO Box 9, Observatory 7935, South Africa
[55] Instituto de Telecomunicações, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
[56] Departamento de Física, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
[57] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
[58] Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK
[59] Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
[60] INAF Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy
[61] Department of Astronomy, University of Cape Town, 7701 Rondebosch, South Africa

Appendix A: Cyclic processing and catalogue changes

Cyclic processing is a reprocessing of all the *Gaia* data that occurs between IDT processing and the main data releases. It includes some of the same kinds of processing as IDT (but there are also many additional activities), but at greater leisure and with more computationally-expensive algorithms. In particular, cyclic processing repeats and revises the mapping of transits to sources, using an improved spatial cross-match with improved astrometry (see Torra et al. 2021). Cyclic processing happens long after GSA transients have been raised and published from a given set of observations. The published alerts are not revised to reflect changes to flux measurement.

Of necessity, alerts processing uses IDT's working source-catalogue to build the lightcurves. Periodically (on exactly two occasions up to the end of 2020) IDT changes its working catalogue to the most-recent cyclic catalogue. To accommodate this, the alerts database has to be updated to the new catalogue, such that old and new observations may still be combined into lightcurves.

There is never a one-to-one mapping between the old and new working catalogues: cyclic processing uses an improved astrometric solution, which can result in splits and mergers of IDT's sources to best fit a new clustering analysis of the sky (see Torra et al. 2021). And importantly for GSA, a magnitude criterion was employed in the clustering algorithm to help disentangle valid and spurious detections into different clusters. This means that transits forming a lightcurve of a published alert could end up assigned to different sources in the new catalogue, especially for a highly variable (e.g. transient) source, where the bright and faint parts of a lightcurve may appear in the catalogue as distinct sources. An additional algorithm was later added to the cyclic processing in an attempt to improve matching for these variable sources.

When IDT's working catalogue is updated, then we map the transits of published alerts to their new sourceIds (if they have changed). Where the update results in multiple sourceIds associated with a single alert, the alert lightcurve is visually inspected including all positionally coincident transits, and if necessary additional sourceIds are assigned to the published alert (these are called *mixed-in sources*). In this event the published lightcurve will be updated with future transits belonging to any of the sources assigned to the published alert.

Appendix B: Photometric one day calibration

In GSA we store raw IDT fluxes in the database, and calibrate all transit photometry on-the-fly via a database function. The parameters used by the function are derived from the Photometric One Day Calibration (PODC). PODC is generated on an approximately monthly timescale by DPCI, but with a one-day cadence. Thus the GSA photometric calibration is always out-of-date, and relies on the generally stable and slowly varying throughput of the *Gaia* instruments. The PODC calibration is a simplified and non-iterative version of the calibration applied by DPCI for the production of the main *Gaia* data releases. We also note that PODC does not include a *link* calibration (Evans et al. 2018) to bring the distinct *Gaia* instrument configurations (gates and window classes) into agreement. Finally, although the PODC calibration includes colour terms, we decided to neglect these in implementation. This is because there are significant numbers of transits of alerting sources where we are missing BP and/or RP photometry (sometimes these data are delivered later). In Fig. B.1 we show the time evolution of the PODC derived calibration factors for the 110 calibration units corresponding to the ungated windows for the faintest sources (fainter than $G_{\rm mag} \sim 16$).

We compared the PODC calibration to the iterative *Gaia* DR2 *G*-band calibration (Evans et al. 2018; Riello et al. 2018) for a set of 184 000 sources, which have been randomly selected to lie in the SDSS DR7 footprint (this avoids the most crowded regions of the Galactic plane). We required the sources to have a minimum of 10 *Gaia* field-of-view transits, and we use the median of the per-CCD PODC calibrated fluxes as representative of the per-transit CCD flux. Because PODC was initialised twice, with two distinct magnitude zeropoints, we apply these internal zeropoints to transform from fluxes to magnitudes. For the DR2 fluxes we use the revised photometric zeropoints[15] published in March 2018.

Some striking features can be seen in the difference between the PODC and the DR2 photometry (see Fig. B.2). The first is that there are two discontinuities (at the few per cent level) at $G({\rm DR2}) \sim 13$ and $G({\rm DR2}) \sim 16$. These both correspond to changes in the on-board window class allocated to a source, thus the size and binning of the readout window (de Bruijne et al. 2015). At $G = 13$ the window changes from 2D (for brighter sources) to 1D (window Class = 1), while at $G = 16$ the 1D window changes size from 18 to 12 pixels Along Scan (window Class = 2). These steps are best explained by the lack of a *link* calibration in PODC (see earlier).

The second feature visible in Fig. B.2 is a clear non-linear trend between offset and magnitude for sources fainter than $G = 16$. The model overlaid on the figure is a fit to the median offset between PODC and DR2 in discrete magnitude bins, ranging from +0.05 at $G = 16.2$ to −0.025 at $G = 20.4$. We use a quadratic model fitted to the magnitude term:

$$G_{\rm PODC} - G_{\rm DR2} = -0.69 + 0.094 G_{\rm DR2} - 0.003 G_{\rm DR2}^2 \tag{B.1}$$

shown as a dashed red line in the same figure. It is not clear where this apparent non-linearity arises, but we note that *Gaia* DR2 takes Intermediate Data Update fluxes as input, which have improved image parameter determination upon that implemented in IDT, as well as a better PSF/LSF model, better background treatment and other improvements (see Lindegren et al. 2018).

The third feature of note in Fig. B.2 is the large scatter in $G_{\rm PODC} - G_{\rm DR2}$ at any $G({\rm DR2})$, and the sequence actually looks bimodal beyond $G({\rm DR2}) \sim 17$. In order to test the hypothesis that this is a colour effect (because we excluded the colour-term in our implementation of PODC), we show the residuals from the model for sources fainter than $G = 16$ in Fig. B.3. Indeed, there is a significant linear trend with BP–RP which explains most of the observed scatter. For completeness, the model fit to the data shown in the figure leads to an updated version of Eq. (B.1), such that:

$$G_{\rm PODC} - G_{\rm DR2} = -0.67 + 0.094 G_{\rm DR2} - 0.003 G_{\rm DR2}^2 - 0.015(BP - RP) \tag{B.2}$$

---

[15] `https://www.cosmos.esa.int/web/gaia/iow_20180316`

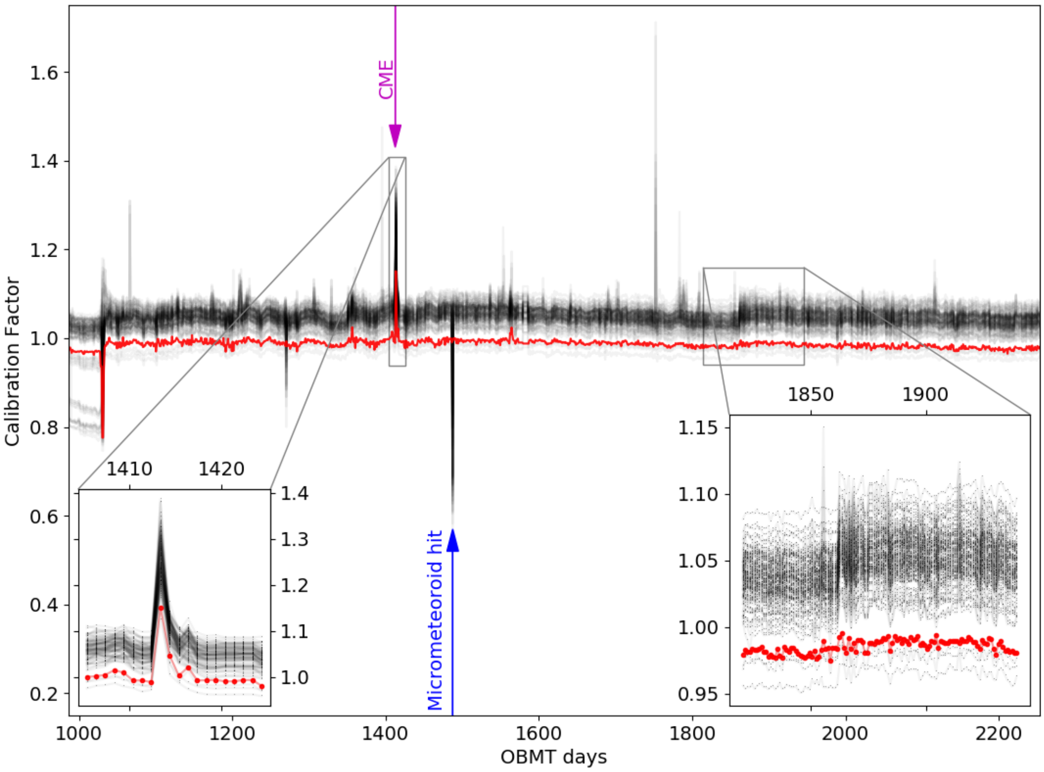

Fig. B.1. Calibration factors returned by PODC for the 110 calibration units corresponding to the ungated windows for the faintest sources (fainter than $G_{\rm mag} \sim 16$). The calibration factors for the calibration unit corresponding to AF8, row 2 in the following field-of-view (FFoV) is shown in red. The time interval covers the range of IDT runs from 1046 to 4724 used in this paper, and is expressed in days in on-board mission time (OBMT). Excursions in the daily measured calibration factor can and do occur, these can result from large numbers of cosmic rays/charged particle events as well as micrometeroid hits. Insert lower left: (indicated by magenta arrow) excursion which occurred roughly 2 days after the CME X9.3 of September 2017. Insert lower right: close-up of the more typical stable behaviour of the calibration factor derived by PODC. We note the jump around OBMT 1861.9 is a refocusing of both FOVs.

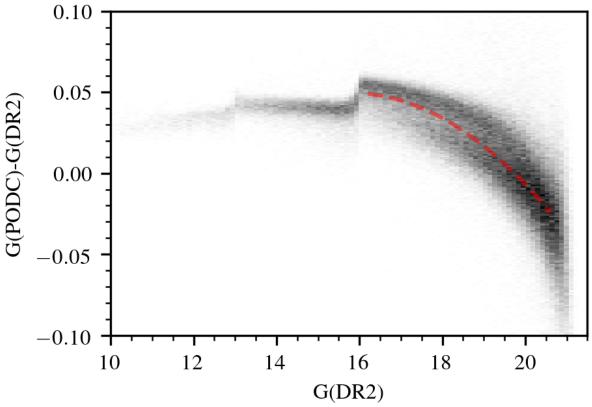

Fig. B.2. Offset between PODC and DR2 calibrated sources as a function of DR2 G magnitude. Sources must have a minimum of 10 *Gaia* transits for inclusion, and were selected to have overlaps with the SDSS footprint, hence avoiding regions of low Galactic latitude and high extinction. A simple quadratic model (shown as a dashed red line) shows the residual magnitude dependence for sources fainter than $G = 16$.

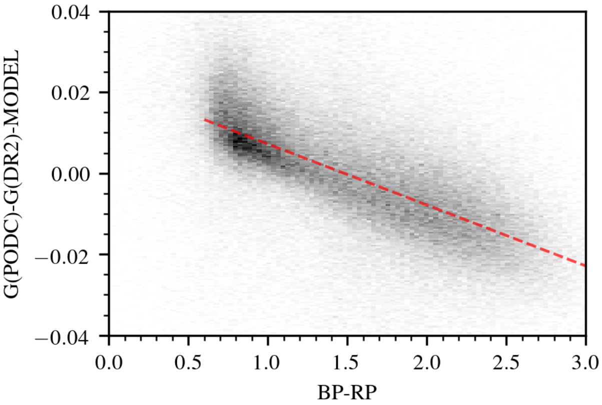

Fig. B.3. Difference between data and quadratic magnitude model (see Fig. B.2) as a function of *Gaia* BP–RP colour.

Appendix C: Computing arrangements

The alerts cluster is co-located with, but distinct from, DPCI. Our cluster is built around a large, PostgreSQL Relational Database Management System (RDBMS). A dedicated, 32-core server holds the primary copies of the databases and an identical machine holds a secondary, read-only copy of each database, kept current by synchronous replication. These replica databases were originally installed as protection against data loss, a role that has never been exercised. They now serve to increase performance by spreading the query load across two servers. Each database server has 176 TB of bulk-data space, arranged as a RAID 6. The alerts computers are interconnected, and connected to DPCI, by an Infiniband network.

The alerts pipeline is divided into: a batch application for ingestion of data, and primary data-reduction; a batch application for filtering of candidate alerts; a web application for human assessment of candidates surviving automatic filtering; and further web-applications for publishing selected alerts. The data reduction application, which handles the greatest volume of data in each run, is assigned its own 64-core server. An identical server is held in readiness to take over should the first one fail. In practice, this reserve server is used for testing, for exploration of new algorithms, and occasionally for large-scale rearrangements of the main database. Each data-reduction server has 22 TB of data space to receive new data from DPCI. This holds roughly 6 months of data and allows reprocessing to correct errors or improve results over this span of the mission.

The other applications run in virtual machines hosted on a pair of smaller servers. There is also a disc-server for back-ups of the main databases, with a capacity of 284 TB.

The main data-reduction application was written in Java, for performance and to conform to DPAC standards. The web applications were written in Python, using the Django framework, for ease of development. The batch-filtering application was also written in Python.

The most notable feature of the computing design is that the bulk of the data is stored in the PostgreSQL database. Compare this with the more common arrangement where only metadata are stored in the RDBMS, while the bulk is stored in flat files. Ingesting both the full input data-set, and the results of processing, reduces significantly the amount of code needed to simply move data, and allows extreme flexibility in writing the applications that consume, analyse and display the results. However, it is challenging to obtain sufficient throughput from the relational architecture. Many design choices in the database and applications are forced by the need to reduce a day's data in six hours or less.

At the time of writing, the PostgreSQL system occupies 110 TB of its storage (with some unknown and variable fraction of this reclaimable for new data), almost all of which space is taken up in ingesting and recording the transit data. This is exceptionally large for a database that is extended daily and in which roughly 5% of the contents must be scanned for daily processing.

The data-set is far too large to fit comfortably into a single table for each record type, and is therefore partitioned by position of the sources into level-5 ($n_{\text{side}} = 32$) HEALpixels; there are 12 288 tables for each record type that deals with individual transits or sources. This partitioning is much coarser than the level-12 mesh used elsewhere by DPAC to group *Gaia* sources into catalogues. The level-5 mesh was chosen because it is the closest match to the width of the *Gaia* field of view. Empirically, it is most efficient to query historic transit-data for a whole HEALpixel, selecting transits only for those sources that have received new data in current run, than to make one query per source with new data. The level-5 mesh minimises simultaneously the number of these expensive queries and the number of rows touched by a query that relate to sources outside the current scan.

The queries to ingest and to raise data cannot be straightforward in such a highly-stressed system; they have to be highly tuned for performance. This precludes the use of generic, object-relational-mapping libraries to generate queries. Instead, the application code forms its queries directly. In order to achieve performance, much parallelism is needed, and the number of threads accessing the data has to be chosen carefully. We knew from pre-mission tests that the application was not the bottleneck and could be given sufficient threads of execution to saturate the database engine. Later experiments showed that the database performance, for this application, is limited by CPU power on the database server, not by I/O capacity; the complexity of the queries dominates. Therefore, the ideal arrangement is to have one application thread accessing the database per CPU core on the database server: 32 such threads in our system. The data-reduction servers have 64 CPU cores and to use these we run two second-level threads per HEALpixel to consume the raised data from the memory of the data-reduction server. We feel that we have reached a practical limit in the size of database that can be hosted on a single RDBMS-server. Any major expansion (beyond routine addition of daily data for the remaining mission) or rearrangement of the database would drive us to a multi-server solution with partitioning across computers. This would be massively more expensive and we are happy that we have managed to do the work with single servers. We note that while it is possible to tune the daily data-flow for acceptable performance, whole-database operations are uncomfortably slow. This includes data-model changes that add columns; weekly back-ups of the database; and especially changes to the IDT working-catalogue that force us to update the transit-source mapping for the whole data-set. The latter operation requires down-time of at least one week. When using PostgreSQL at this scale, any routine operation becomes a significant campaign that requires careful planning.

Appendix D: Abbreviations used in this paper

AC – across-scan direction
ADS – Astrophysics Data System
AF – Astrometric Field
AGIS – Astrometric Global Iterative Solution
AL – along-scan direction
ASAS – All-Sky Automated Survey
ASAS-SN – The ASAS Supernova survey
ATEL – Astronomer's Telegram
BP/RP – Blue Photometer/Red Photometer
CBAT – Central Bureau for Astronomical Telegrams
CCD – charge-coupled device (detector)
CME – Coronal Mass Ejection
CPU – Central Processing Unit
csv – comma separated values (file type)
CV – Cataclysmic Variable
Dec – Declination
DPAC – Data Processing and Analysis Consortium
DPCI – Data Processing Centre in Cambridge
DR2 – *Gaia*'s Data Release 2, 25 April 2018
DR3 – *Gaia*'s Data Release 3, expected 2022
ESA – European Space Agency
ESO – European Southern Observatory
ESOC – European Space Operations Centre (Darmstadt, Germany)
ETL – Extract Transform Load
FOV – Field of View
G – *Gaia*'s 'white light' photometric band
GDR2 – *Gaia* Data Release 2
GoF – Goodness-of-Fit
GSA – *Gaia* Science Alerts
HEALpix – Hierarchical Equal Area isoLatitude Pixelisation
HRD – Hertzsprung-Russell diagram
IAU – International Astronomical Union
IDT – Initial Data Treatment
IGSL – Input *Gaia* Source List
IoA – Institute of Astronomy, Cambridge
IPAC – Infrared Processing and Analysis Center (NASA)
iPTF – intermediate Palomar Transient Factory
LSF – Line Spread Function
MAD – Median Absolute Deviation
MASTER – Mobile Astronomical System of Telescope-Robots
MOC – Mission Operations Centre (Darmstadt, Germany)
NASA – National Aeronautics and Space Administration
NED – NASA/IPAC Extragalactic Database
NUTS – Nordic Optical Telescope Unbiased Transient Survey
OBMT – On Board Mission Time
OGA – On-ground Attitude
OGLE – Optical Gravitational Lensing Experiment
PanSTARRS – Panoramic Survey Telescope and Rapid Response System
PESSTO – Public ESO Spectroscopic Survey of Transient Objects
PODC – Photometric One-day Calibration
PPE – Prompt Particle Event
PSF – Point Spread Function
QSO – Quasi Stellar Object
RA – Right Ascension
RAID – Redundant Array of Independent Discs
RBF – Radial Basis Function
RDBMS – Relational Database Management System
RMS – root mean square
RSS – Really Simple Syndication
RVS – Radial Velocity Spectrograph (*Gaia*)
SDSS – Sloan Digital Sky Survey
SEDM – Spectral Energy Distribution Machine
– (spectrograph)
SLSN – Superluminous Supernova
SM – Sky mapper (*Gaia*)
SN – Supernova
SOC – Science Operations Centre (Madrid, Spain)
SSO – Solar System object
SVM – Support Vector Machine
TCB – Barycentric Coordinate Time
TNS – Transient Name Server
URL – Uniform Resource Locator (web address)
UTC – Coordinated Universal Time
VOEvent – the Virtual Observatory Event language
XP – shorthand for BP/RP
XRB – X-ray binary star
VSX – International Variable Star Index database
YSO – Young Stellar Object
ZTF – Zwicky Transient Facility