Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **481,** 3357–3369 (2018)
Advance Access publication 2018 September 13

doi:10.1093/mnras/sty2498

The *Gaia*/IPHAS and *Gaia*/KIS value-added catalogues

S. Scaringi [1,2]★ C. Knigge,[3] J. E. Drew [4] M. Monguió,[4] E. Breedt [5] M. Fratta,[1,2] B. Gänsicke,[6] T. J. Maccarone,[1] A. F. Pala [6] and C. Schill[2]

[1]*Department of Physics and Astronomy, Texas Tech University, Lubbock, TX 79409-1051, USA*
[2]*School of Physical and Chemical Sciences, University of Canterbury, Christchurch 8041, New Zealand*
[3]*School of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK*
[4]*School of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield, Hertfordshire AL10 9AB, UK*
[5]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[6]*Astronomy and Astrophysics Group, Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK*

Accepted 2018 September 7. Received 2018 September 7; in original form 2018 July 20

ABSTRACT

We present a sub-arcsecond crossmatch of *Gaia* DR2 against the INT Photometric H α Survey of the Northern Galactic Plane Data Release 2 (IPHAS DR2) and the *Kepler*-INT Survey (KIS). The resulting value-added catalogues (VACs) provide additional precise photometry to the *Gaia* photometry (r, i, and H α for IPHAS, with additional U and g for KIS). In building the catalogue, proper motions given in *Gaia* DR2 are wound back to match the epochs of IPHAS DR2, thus ensuring high proper motion objects are appropriately crossmatched. The catalogues contain 7927 224 and 791 071 sources for IPHAS and KIS, respectively. The requirement of $>5\sigma$ parallax detection for every included source means that distances out to 1–1.5 kpc are well covered. We define two additional parameters for each catalogued object: (i) f_c, a magnitude-dependent tracer of the quality of the *Gaia* astrometric fit; (ii) f_{FP}, the false-positive rate for parallax measurements determined from astrometric fits of a given quality at a given magnitude. Selection cuts based on these parameters can be used to clean colour–magnitude and colour–colour diagrams in a controlled and justified manner. We provide both full and light versions of the VAC, with VAC-light containing only objects that represent our recommended trade-off between purity and completeness. Uses of the catalogues include the identification of new variable stars in the matched data sets, and more complete identification of H α-excess emission objects due to separation of high-luminosity stars from the main sequence.

Key words: catalogues – surveys – parallaxes – proper motions – Galaxy: stellar content – stars: emission-line.

1 INTRODUCTION

The European Space Agency's *Gaia* mission provides unprecedented opportunities to assemble reliable Hertzsprung–Russell diagram for different types of stellar populations. *Gaia* Data Release 2 (DR2, Gaia Collaboration et al. 2016, 2018a,b; Arenou et al. 2018; Lindegren et al. 2018) includes photometry in the G, G_{BP}, and G_{RP} bands (see Fig. 1) for approximately 1.5 billion sources. Its quality and size will define the new standard in the years to come, and have a tremendous impact on various areas of astrophysics. In particular, it is the astrometry, and specifically the parallax measurements, that will provide the largest impact, since it is with these measurements that we can now infer distances, absolute magnitudes, and transverse velocities (with the additional proper motion information) for individual targets.

The INT/WFC Photometric H α Survey of the Northern Galactic Plane (IPHAS, Drew et al. 2005) is the first comprehensive digital survey of the northern Galactic disc ($|b| < 5°$), covering a Galactic longitude range of $29° < l < 215°$. The IPHAS observations are obtained using the Wide Field Camera (WFC) at the prime focus of the 2.5 m Isaac Newton Telescope (INT) on La Palma, Spain. IPHAS images are taken through three filters: a narrow-band H α, and two broad-band Sloan r and i filters (see Fig. 1). Exposures are set to reach an r-band depth of $\approx$21. Pipeline data reduction is handled by the Cambridge Astronomical Survey Unit. Further details on the data acquisition and pipeline reduction can be found in Drew et al. (2005) and González-Solares et al. (2008). In this paper we use the IPHAS Data Release 2 (Barentsen et al. 2014), containing measurements for $\approx$219 million sources observed between 2003 and 2012 which have all been photometrically calibrated.

The *Kepler*-INT Survey (KIS, Greiss et al. 2012) observed the *Kepler* field using the same observing strategy as IPHAS on the

★ E-mail: simone.scaringi@ttu.edu

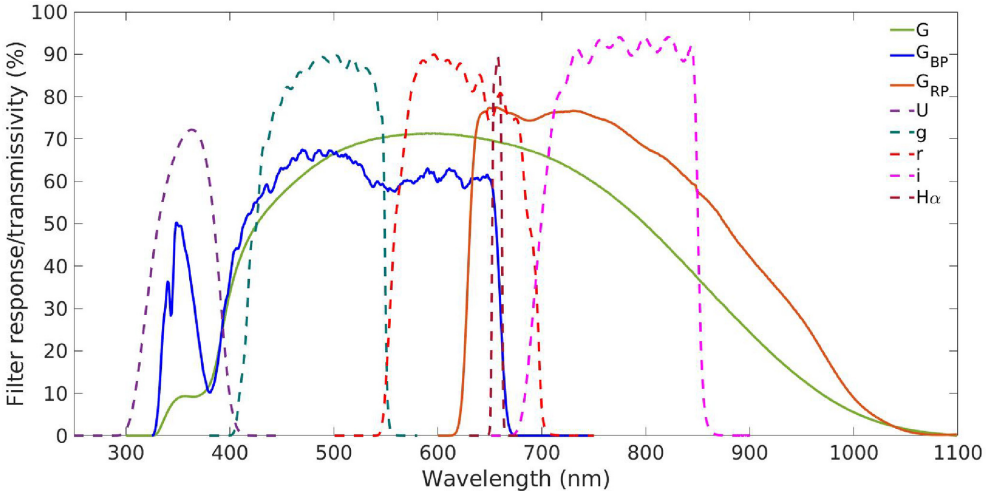

Figure 1. Total response curves from *Gaia* (G, G_{BP}, G_{RP}) and filter transmission curves from IPHAS (r, i, Hα) and KIS (U, g, r, i, Hα).

INT telescope, with additional observations in the Sloan g' band and non-standard U-band (see Fig. 1), both of which are also used in the UV-Excess Survey of the Northern Galactic Plane (Groot et al. 2009). The pipeline data reduction is identical to IPHAS. In this paper we use the KIS Data Release 2, which provides coverage of 97 per cent of the *Kepler* field, and contains ≈14.5 million photometrically calibrated entries.

This paper presents sub-arcsecond crossmatches between the *Gaia*/IPHAS and *Gaia*/KIS catalogues, taking into account the different epochs of observations of both IPHAS and KIS, as well as the proper motion information for each target in the *Gaia* catalogue. In the process of producing these catalogues we additionally include, for each target, two additional columns: a so-called completeness fraction (f_c) which provides information relating to how acceptable the *Gaia* astrometric solution is compared to targets with similar G-band magnitudes, and a so-called false-positive fraction (f_{FP}) providing information on how reliable the astrometric measurements of a given target are.

Section 2 describes our crossmatching procedure, including the preliminary selection cuts applied to all data sets and examples of recovered matches. Section 3 introduces our additional quality control parameters f_c and f_{FP}, and discusses how these can be used to clean the *Gaia*/IPHAS and *Gaia*/KIS catalogues from unreliable entries. Section 4 provides some illustrative examples of how our value-added catalogues (VACs) can be used for science exploitation. Finally Section 5 describes our published catalogue formats, with conclusions drawn in Section 6.

2 CROSSMATCHING *GAIA* WITH IPHAS AND KIS

The *Gaia* DR2 release contains results for over 1.6 billion sources. The majority of these data are not required for our crossmatching purposes since it either lies outside the IPHAS/KIS footprint and/or the *Gaia* results are not of high enough quality. In this section we describe how we performed a sub-arcsecond matching between the IPHAS/KIS targets with *Gaia* DR2, including descriptions of the selection cuts and proper motion corrections. We also highlight the advantage of our method against a simple crossmatch through some examples.

2.1 Selection cuts

Before attempting to crossmatch sources in IPHAS/KIS with *Gaia* DR2, we apply some quality cuts to all data sets in order to retain only sources with good photometric and astrometric measurements.

From IPHAS DR2 we select only objects which:

(i) have measurements in all three bands (r, i, and Hα);
(ii) are fainter than the saturation limit in all three photometric bands ($r > 13$, $i > 12$, and H$\alpha > 12.5$);
(iii) have photometric errors smaller or equal to 0.1 mag in all bands;
(iv) are not flagged as blended or affected by bright neighbours in any band.

Of the 218 991 524 sources in IPHAS DR2 63 520 381 survive these quality cuts. Similar cuts are applied to the KIS DR2 catalogue, with the inclusion of the same cuts in the U and g bands. This retained 2662 117 sources out of 14 476 957.

From *Gaia* DR2 we select only objects which:

(i) have a G-band flux signal-to-noise above 5 (`phot_g_mean_flux_over_error`>5);
(ii) have a signal-to-noise parallax measurement above 5 (`parallax_over_error`>5);
(iii) are within an area slightly larger than the IPHAS footprint ($20 < l < 220$ and $-6 < b < 6$);
(iv) are within the KIS footprint ($275 < \alpha_{J2015.5} < 305$ and $36 < \delta_{J2015.5} < 54$).

The above selection criteria yield two *Gaia* data sets: one containing 19 553 253 sources within the IPHAS footprint, and one containing 3004 331 sources in the KIS footprint.

2.2 Proper motion corrections and crossmatching

In order to minimize mismatches between the *Gaia* catalogue and both IPHAS and KIS, as well as recovering fast-moving objects, it is important to take into account the proper motion of targets. *Gaia* DR2 provides proper motion information for all targets within the IPHAS and KIS footprints which pass our data quality cuts. However, given the way they were designed, neither IPHAS and KIS contain this information. Furthermore, although all catalogues

give positions in the barycentric ICRS reference frame, only the *Gaia* DR2 positions are given at epoch 2015.5. Both IPHAS and KIS report positions at the epoch of observation, which can be any time between 2003 and 2012 for IPHAS and between 2011 and 2012 for KIS. The epoch of observation for a particular target is reported in both IPHAS and KIS DR2 catalogues.

Ideally, for precise crossmatching between the catalogues, the *Gaia* astrometry would have to be propagated to the IPHAS/KIS epoch of each source before the crossmatching is performed. This would result in recomputing the *Gaia* astrometry for every entry in the input catalogues (in excess of 20 million when considering both IPHAS and KIS), and becomes even more unfeasible for larger input catalogues. Instead we proceed by first dividing the IPHAS- and KIS-cleaned catalogues into monthly epoch batches. Because of the observing strategy of both IPHAS and KIS, which sequentially observe all bands immediately following each other, we take the epoch of a particular target to be the start of the *r*-band observation. This ensures that the epoch-corrected positional uncertainty of the *Gaia* catalogue is relativity small even for high proper motion objects. For example, the recomputed *Gaia* coordinates for an object with an extreme proper motion of 2 arcsec year^{-1} will be at worst ≈0.08 arcsec off the IPHAS/KIS position.

This procedure results in 46 monthly batches for IPHAS and six for KIS. For each of these batches, we then recompute the *Gaia* astrometry to the mid-point epoch for each month. We then select the positional closest match in the sky within 1 arcsec of a given IPHAS or KIS entry. After removing for *Gaia*-duplicated sources,[1] this retains 7927 532 and 827 989 sources for the IPHAS and KIS footprints, respectively. However, because some areas of the sky in both IPHAS and KIS have been observed more than once (excluding offset fields), some of the retained sources will have duplicated entries in our catalogues. We thus clean the retained sources by removing duplicates based on their *Gaia* DR2 designation. The retained number of sources is then 7927 224 and 791 071 for IPHAS and KIS, respectively.

To ensure that the correct match is found in cases where two or more targets are within the 1 arcsec crossmatch radius, we retained all matches found within 1 arcsec when crossmatching IPHAS to *Gaia*. In total there are 3253 pairs (no triples or more) which can be found within 1 arcsec when doing the crossmatch. We chose to then inspect the $G_{\rm RP} - i$ colour for these to determine whether this information could help reduce any false-positive matches. Although the *Gaia* $G_{\rm RP}$ fluxes are derived from simple integration of 3.5 by 2.1 arcsec windows (and thus cannot resolve ambiguous matches), the $G_{\rm RP}$ and i combination has been chosen since the i-band is the only IPHAS band to fully reside within a *Gaia* band (see Fig. 1). Of the 3253 pairs, only four targets have an absolute $G_{RP} - i$ value >1 mag. In all four cases the correct match was identified as being the closer target. Visual inspection of the remaining targets also reveals the closest match to be the correct one.

2.3 Recovered matches and removed false-positives

In order to investigate the efficiency of correcting for proper motion in epochs of monthly batches, we have crossmatched the cleaned IPHAS and KIS catalogues to the cleaned *Gaia* catalogue at epoch J2000, selecting the closest match within a generous 5 arcsec radius for every input target. This exercise mimics what would happen if one used the CDS XMatch service[2] for crossmatching IPHAS DR2 and *Gaia* DR2. By comparing the results of this 'raw' crossmatch to the catalogue produced by our proper-motion corrected crossmatch, we can identify false-positives, false negatives, and mismatches in the 'raw' catalogues.

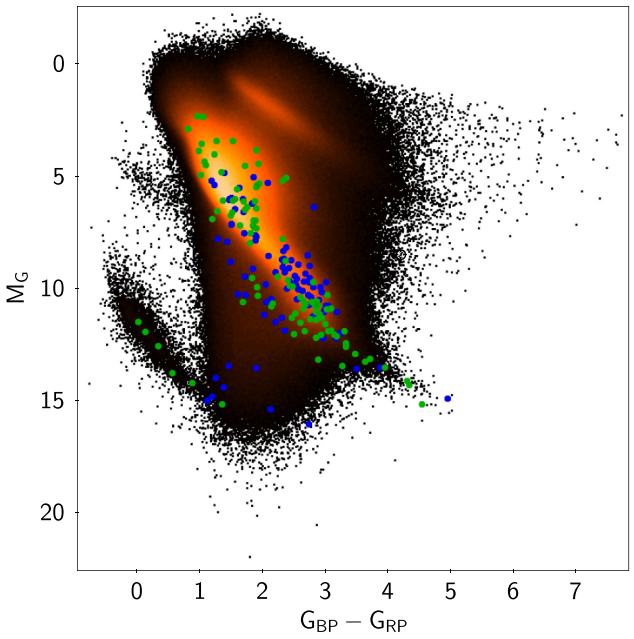

Figure 2. CMD from our crossmatched *Gaia*/IPHAS catalogue using *Gaia* photometry, adopting distances inferred from the *Gaia* parallaxes. The figure also displays objects that are only correctly recovered through the additional *Gaia* proper motion information. Objects which would have been entirely missed are marked in blue, while mismatches are marked in green.

Fig. 2 shows some colour–magnitude diagrams (CMD) for the retained sources within the IPHAS footprint. We note that all distances inferred in this paper have been determined via $M = m + 5 + 5\log_{10}(\varpi/1000)$, where M and m are the absolute and apparent magnitudes, respectively, and ϖ the parallax in milliarcseconds (the same practice as presented by Gaia Collaboration et al. 2018b). These have been computed to generate the CMDs, and are not used for any selection or crossmatching purposes. We have also computed distances using the Exponentially Decreasing Space Density approach (Bailer-Jones et al. 2018; Luri et al. 2018), adopting a scale height of $L = 1.35$ kpc, but the results are qualitatively similar. We point out that M and colour for all objects in Fig. 2, as well as all other CMDs plotted in this paper, have not been corrected for extinction, and hence are upper limits on true absolute magnitude and colour.

We recover 103 sources which would have been entirely missed by employing a simple 5 arcsec search radius without proper motion correction (blue circles), while 101 sources would have been mismatched (green circles). More importantly, 209 307 false-positive sources would have erroneously been included. The location of these mismatched sources in the CMD are shown in Fig. 3. Although the *Gaia*-based CMD shows some targets on the main sequence, the IPHAS-based CMD clearly shows that these false-positives do not lie on any known sequence (due to the erroneous match between the

[1]Targets flagged as duplicate sources in the *Gaia* archive may indicate observational, cross matching or processing problems, or stellar multiplicity, and probable astrometric or photometric problems in all cases.

[2]http://cdsxmatch.u-strasbg.fr/xmatch

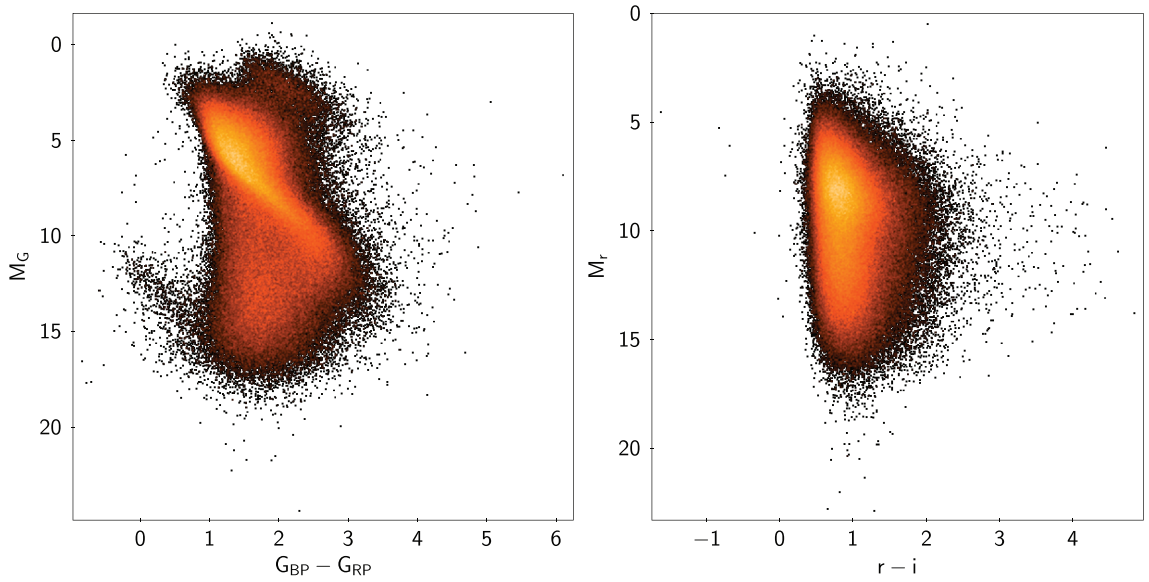

Figure 3. CMDs for IPHAS objects which would have been wrongly been associated to a *Gaia* counterpart using a standard 5 arcsec crossmatch radius, and ignoring proper motion information. The left-hand panel displays the *Gaia* CMD, while the right-hand panel displays the corresponding IPHAS CMD for the same sources.

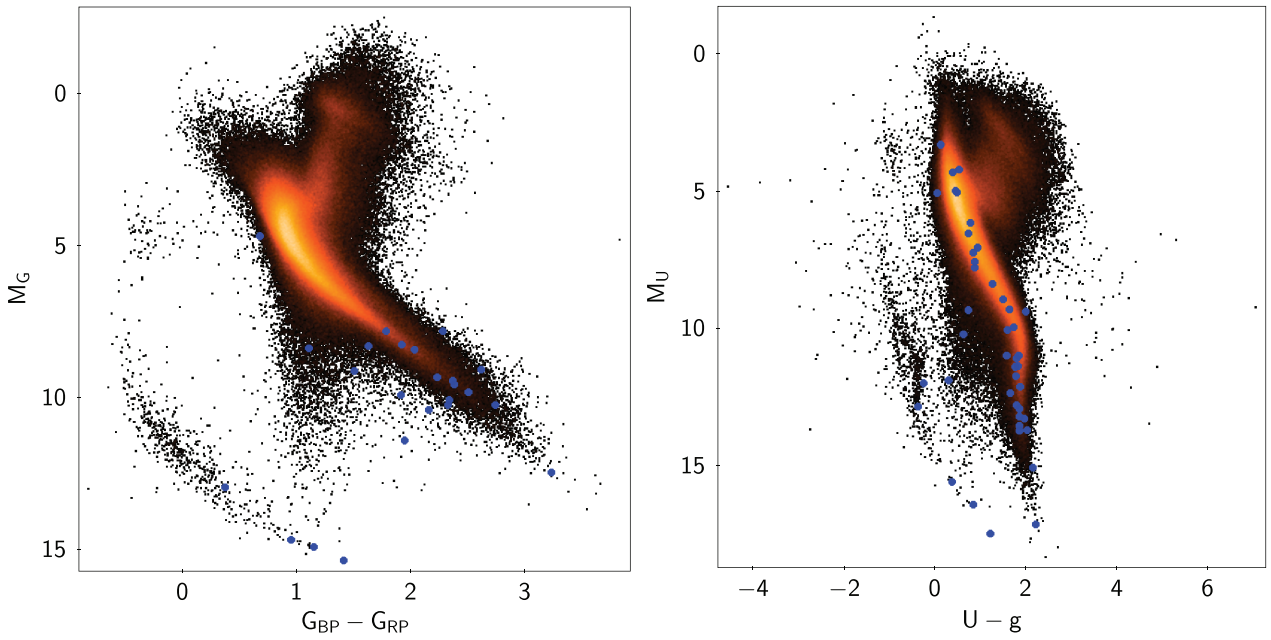

Figure 4. CMDs from our crossmatched *Gaia*/KIS catalogue using *Gaia* photometry (left-hand panel) and KIS photometry (*U* and *g*, right-hand panel), both adopting distances inferred from the *Gaia* parallaxes. Both panels also display objects which would have been entirely missed without correcting for proper motion.

IPHAS colours and the *Gaia* distances), but rather occupy a region in parameter space which is known to be populated by stars with problematic distance estimates (Lindegren et al. 2018).

Fig. 4 shows some CMDs for our matched targets in the KIS footprint, including 43 targets which would have been missed without taking proper motion to account (blue circles). It is worth noting that the inclusion of the *U* and *g* bands from KIS clearly separates the H and He white dwarf tracks, as has been shown with SDSS (Sloan Digital Sky Survey) colours in Gaia Collaboration et al. (2018b). We have omitted displaying 35 639 targets which have been mismatched for clarity, but these show qualitatively similar problems as those shown in Fig. 3 from the IPHAS sample.

It is interesting that the number of mismatches, as well as the number of duplicates (36 184), is relatively higher in the uncorrected *Gaia*/KIS catalogue than in the uncorrected *Gaia*/IPHAS one. This is due to the way the *Kepler* footprint was tiled in KIS, with some areas being observed multiple times. This is illustrated in Fig. 5, which shows the sky position of the *Gaia*/KIS catalogue

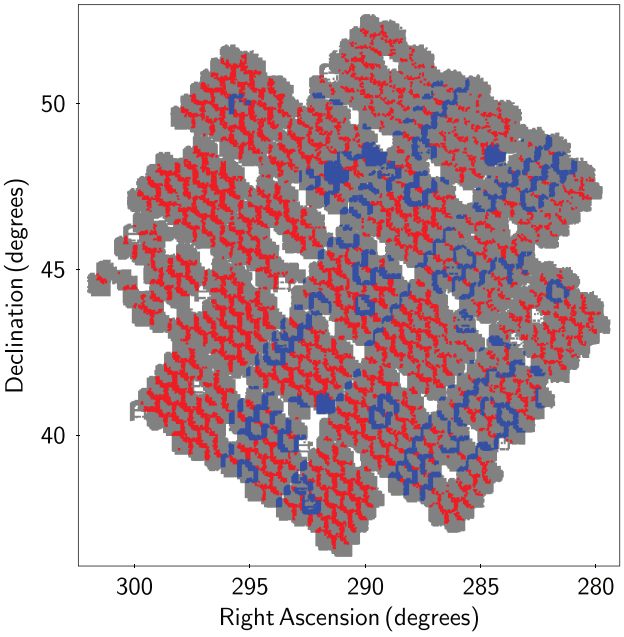

Figure 5. Sky position for all targets retained in our *Gaia*/KIS catalogue (grey points). Overlayed are targets which would have been mismatched using a standard 5 arcsec crossmatch radius (red points), as well as targets which appeared as duplicates in our final *Gaia*/KIS catalogue due to being observed multiple times during the KIS survey (blue points).

with the mismatches and duplicate sources displayed in red and blue, respectively. The outline of the overlap regions is clearly visible.

2.4 Bias induced by selection cuts

The selection cuts described in Section 2.1 will introduce a number of mismatches between the IPHAS/KIS catalogues and the epoch-corrected *Gaia* catalogue. In particular, the selection on `phot_g_mean_flux_over_error`>5 will have a large effect, since it leads to many real *Gaia* detections not being available for crossmatching. As a result, a given IPHAS/KIS target may be matched to the wrong *Gaia* counterpart, as the true counterpart might have not passed our initial cuts. To estimate how large this effect is in our catalogues we reran our crossmatching algorithm on the full *Gaia* DR2 catalogue, without imposing any selection cuts. In order to keep this task manageable, we performed this exercise only in a moderately dense region of the IPHAS footprint (Farnhill et al. 2016), with $60 < l < 70$. Estimating the mismatch fraction in this region will then yield an upper limit on the mismatch fraction throughout both the *Gaia*/IPHAS and *Gaia*/KIS catalogues.

In total, there are over 27 million *Gaia* targets within $60 < l < 70$ and $-6 < b < 6$. Of these, ≈6 million entries can be matched to an IPHAS source within 1 arcsec. Our *Gaia*/IPHAS catalogue contains 624 117 within the same footprint area, and we find 726 sources to have been mismatched based on their *Gaia* DR2 designation. We can thus place an upper limit of 0.1 per cent on the fraction of mismatches associated with our selection cuts, and note that many of these mismatches are very faint ($G > 19$) *Gaia* sources, and that this effect will be much lower for the *Gaia*/KIS catalogue given the lower crowding of sources above the Galactic plane.

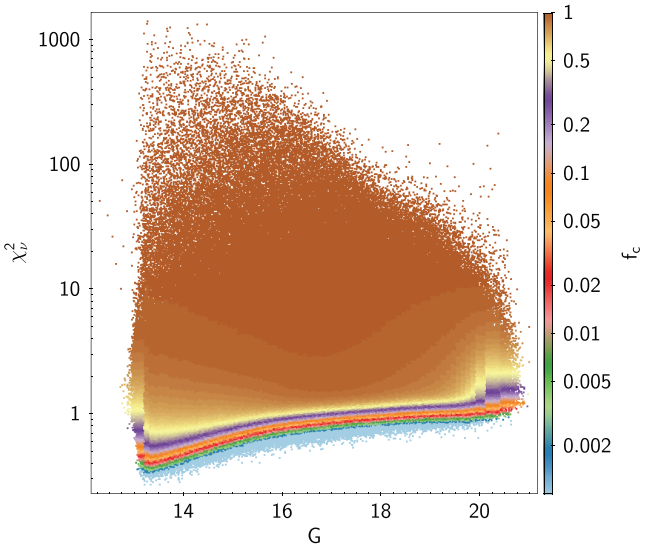

Figure 6. *Gaia* *G*-band magnitude versus reduced χ^2 (χ^2_ν, based on the *Gaia* astrometric fit) for all targets retained in our *Gaia*/IPHAS catalogue. Targets are colour coded by their 'completeness' fraction f_c (see Section 3.1 for details).

3 PURITY VERSUS COMPLETENESS: INTRODUCING ADDITIONAL QUALITY PARAMETERS

In this section we will introduce some additional quality parameters which can be used to clean our merged *Gaia*/IPHAS and *Gaia*/KIS catalogues from targets with unreliable parallax measurements.

Even though we will only explicitly discuss the *Gaia*/IPHAS crossmatch below, the same procedure has also been applied to the *Gaia*/KIS catalogue, with similar results.

3.1 Completeness

Ideally, sources from the *Gaia* catalogue that have poor astrometric solutions can be removed using the goodness-of-fit statistic provided by the *Gaia* Archive (see discussion in Lindegren et al. 2018). Rather than remove targets, we have opted to retain all sources, and instead include for each source a 'completeness' value representing how good/bad the astrometric fit of a particular target is compared to targets within a similar apparent magnitude range (We will explain the reason for this choice of terminology below).

To do this we first recomputed, for each target, the reduced χ^2 as $\chi^2_\nu =$ `astrometric_chi2_al/astrometric_n_good_obs_al` − 5. We then binned all sources in increasing *G*-band bins of 0.1 mag, with the requirement that each bin contains at least 10 000 sources. Each source was then assigned a percentile based on its χ^2_ν within their corresponding *G*-band magnitude bin. We refer to this percentile as the 'completeness fraction', f_c, because it permits the targeted removal of sources with poor astrometry while retaining any desired completeness fraction. For example, removing all sources with $f_c > 0.9$ removes exactly 10 per cent of all sources (for a completeness of 90 per cent). Fig. 6 shows the resulting (G, χ^2_ν) plane for targets in our *Gaia*/IPHAS catalogue colour coded with f_c. The apparent hard edges are the result of our binning scheme, and could be improved with larger number of objects.

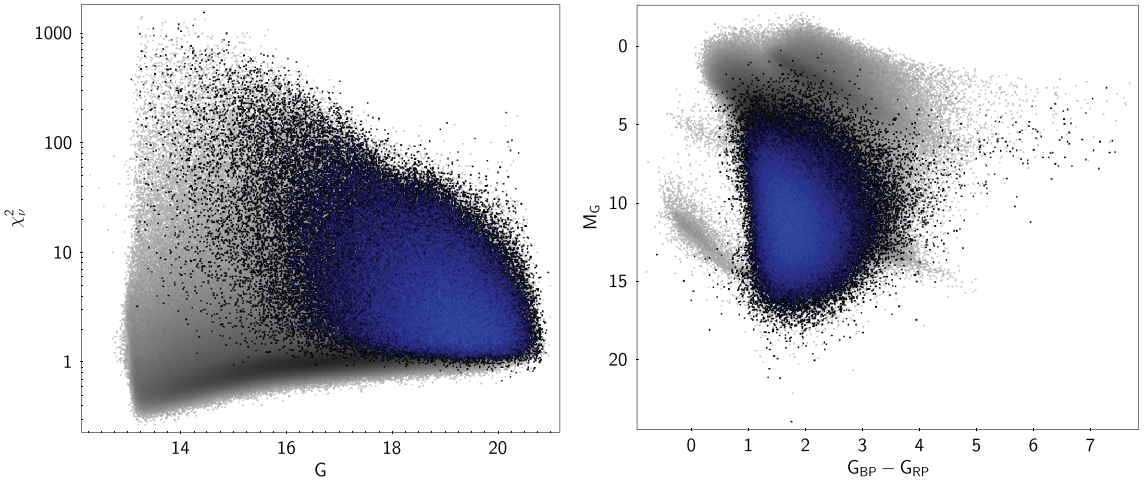

Figure 7. The actual *Gaia*/IPHAS catalogue is shown with grey points, while our negative parallax 'mirror sample' – which highlights the regions where we can expect false-positive to occur – is shown in blue/black. Left-hand panel: *G*-band magnitude versus reduced χ^2 (χ^2_ν, based on the *Gaia* astrometric fit). Right-hand panel: *Gaia*-based CMD for the same targets. Although not realistic, we have adopted absolute values of the parallax measurements to infer the absolute magnitudes of the 'mirror' negative parallax sample to compare their position in the CMD to those 'problematic sources' (Lindegren et al. 2018).

3.2 Purity

It is known and well documented (Lindegren et al. 2018) that *Gaia* DR2 contains some spurious measurements, most notably very large or negative parallaxes. These spurious results are usually traced back to internal crossmatching issues or resolved (or partially resolved) binaries, and are expected to be corrected for in future *Gaia* releases (Lindegren et al. 2018). To mitigate this, Lindegren et al. (2018) have inspected where targets with negative parallaxes fall in *G*-magnitude versus u parameter space, where $u = \sqrt{\chi^2_\nu}$. They then used this to define a simple threshold cut in (G, u) space that is designed to preferentially remove spurious measurements.

Following a similar procedure to Lindegren et al. (2018) we have produced a 'mirror sample' of our crossmatch catalogue that includes *only* sources with spurious *Gaia* measurements that are nevertheless as 'convincing' as those in our actual catalogue. In order to construct this mirror sample, we query the *Gaia* DR2 archive with the same criteria described in Section 2.1, except we change the `parallax_over_error>5` condition to `parallax_over_error<−5`. This query resulted in 603 742 sources for IPHAS and 25 295 sources for KIS. We then parsed our mirror sample through the same crossmatching procedure described in Section 2.2. This allows us to have a sample for which we know the astrometry is bad, but still retains the statistical properties of our catalogues.

The left-hand panel of Fig. 7 shows the (G, χ^2_ν) plane for targets in our catalogue, together with our negative parallax mirror sample overlayed. The right-hand panel shows the resulting *Gaia* CMD for the full catalogue with our mirror sample overlayed, where we have used the absolute parallax value to infer absolute magnitude. The assumption made in adopting absolute parallax values for negative parallax targets is that the *Gaia* DR2 processing occasionally produces spurious astrometry that may equally well result in a positive or negative parallax. This assumption is tested using the absolute value. The similarity in the location of the 'blob' in CMD space to where most targets are expected not to be reliable is remarkable. Also striking is the similarity between the region in CMD space occupied by our mismatched sample (Fig. 3, right-hand panel) and our mirror sample, confirming our assumption.

Using both our catalogue and the mirror sample we define a new quality parameter for each object in our catalogue which identifies the probability for a particular object to be a false-positive entry based on its position in the (G, χ^2_ν) plane. We do this by first binning all targets (including our mirror sample) in *G*-band bins of 0.1 mag, with the additional requirement that each bin contains at least 10 000 objects, similar to what has been done in Section 3.1. For each *G*-band bin, we then sort all entries by increasing χ^2_ν, and subsequently bin these into blocks of 1000 objects. For each target in our catalogue which lies in a particular block we then define the false-positive fraction ($f_{\rm FP}$) as the number of objects in the mirror sample in this block, divided by the total number of objects in the block (i.e. $f_{\rm FP} = \frac{N_{\rm neg}}{N_{\rm pos}+N_{\rm neg}}$, where $N_{\rm neg}$ and $N_{\rm pos}$ are the number of negative and positive parallax objects within a specific block). It is important to realize that our definition of $f_{\rm FP}$ ensures that the obtained values will strictly be within the $0 \leq f_{\rm FP} < 1$ range. In the low $f_{\rm FP}$ regime, applicable to the *Gaia* data set, our definition can be interpreted as $f_{\rm FP} = \frac{N_{\rm neg}}{N_{\rm pos}+N_{\rm neg}} \approx \frac{N_{\rm neg}}{N_{\rm pos}}$.

Fig. 8 shows the (G, χ^2_ν) plane colour-coded with $f_{\rm FP}$. Similar to Fig. 6, the apparent 'edges' in $f_{\rm FP}$ are caused by the set limits on the bin/block sizes and/or number of objects per bin/block. This method can potentially be applied to other larger catalogues where this effect can be minimized with smaller bins/blocks. Furthermore, given the requirement of 1000 objects per block, the precision with which we can determine $f_{\rm FP}$ is limited to 0.001.

3.3 Cleaning the *Gaia*/IPHAS and *Gaia*/KIS CMDs

Having included for each target in our catalogues two additional quality parameters, $f_{\rm c}$ and $f_{\rm FP}$, we can now use these to clean the catalogues to produce more reliable sets of targets.

Arenou et al. (2018) show that spurious astrometric solutions are more common in specific areas of the sky, and may depend on both the *Gaia* scan directions and epoch of observation. Additionally, spurious astrometric solutions are more frequent in dense areas of the sky, particularly the Galactic plane. Our approach of computing $f_{\rm c}$ and $f_{\rm FP}$ is independent of sky position, and it may be that these values differ depending on sky position. However, as IPHAS

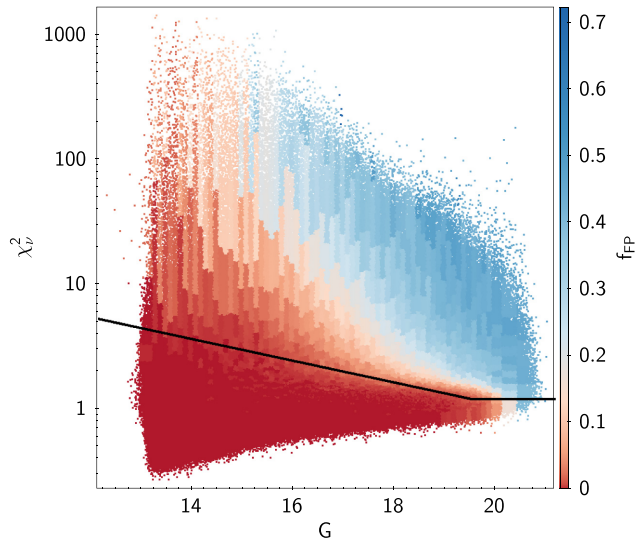

Figure 8. *Gaia* G-band magnitude versus reduced χ^2 (χ^2_ν, based on the *Gaia* astrometric fit) for all targets retained in our *Gaia*/IPHAS catalogue. Targets are colour coded by their false-positive rate fraction f_{FP}, based on the number density of targets with negative parallaxes occupying a specific region in the (G, χ^2_ν) parameter space (see Section 3.1 for details). The thick black line shows the recommended cut discussed in the Appendix of Lindegren et al. (2018) (equation C1).

is concentrated on the Galactic plane where most of the *Gaia* spurious measurements are found, these differences should be relatively small. We also expect the f_{FP} fraction to become more relaxed for other sky areas out of the Galactic plane.

Ideally, if the only problem with our catalogues was the sort of statistical error that is responsible for the existence of the mirror negative parallax sample, there would be no need to have the additional f_c parameter. However, in practice, the sole use of f_{FP} does not remove all sources with bad astrometry. It is therefore useful to use both f_c and f_{FP} to clean our catalogues. We note that when both quality cuts are employed, the 'completeness' of a subset can no longer be guaranteed to be greater than f_c, since it might occur that within specific G-band magnitude bins the f_{FP} threshold will remove additional objects.

To illustrate the effect of selecting targets based on our quality parameters, we show in Figs 9 and 10 the *Gaia* CMDs for varying f_c and f_{FP} thresholds, respectively. The top panels in both figures show the retained targets, while the bottom panels show the removed ones.

Based on visually inspecting various CMDs with all of the *Gaia*, IPHAS, and KIS photometry, our recommended general-purpose quality cuts are $f_c < 0.98$ and $f_{FP} \leq 0.02$. Using these thresholds retains 94 per cent and 98 per cent from our *Gaia*/IPHAS catalogue and *Gaia*/KIS catalogues, respectively. These cuts provide a relatively clean subset of objects, with minimal false-positives and a large retention fraction. Fig. 11 shows some of the CMDs and colour–colour plots produced using our recommended quality cuts for both the IPHAS and KIS merged-catalogues with *Gaia*.

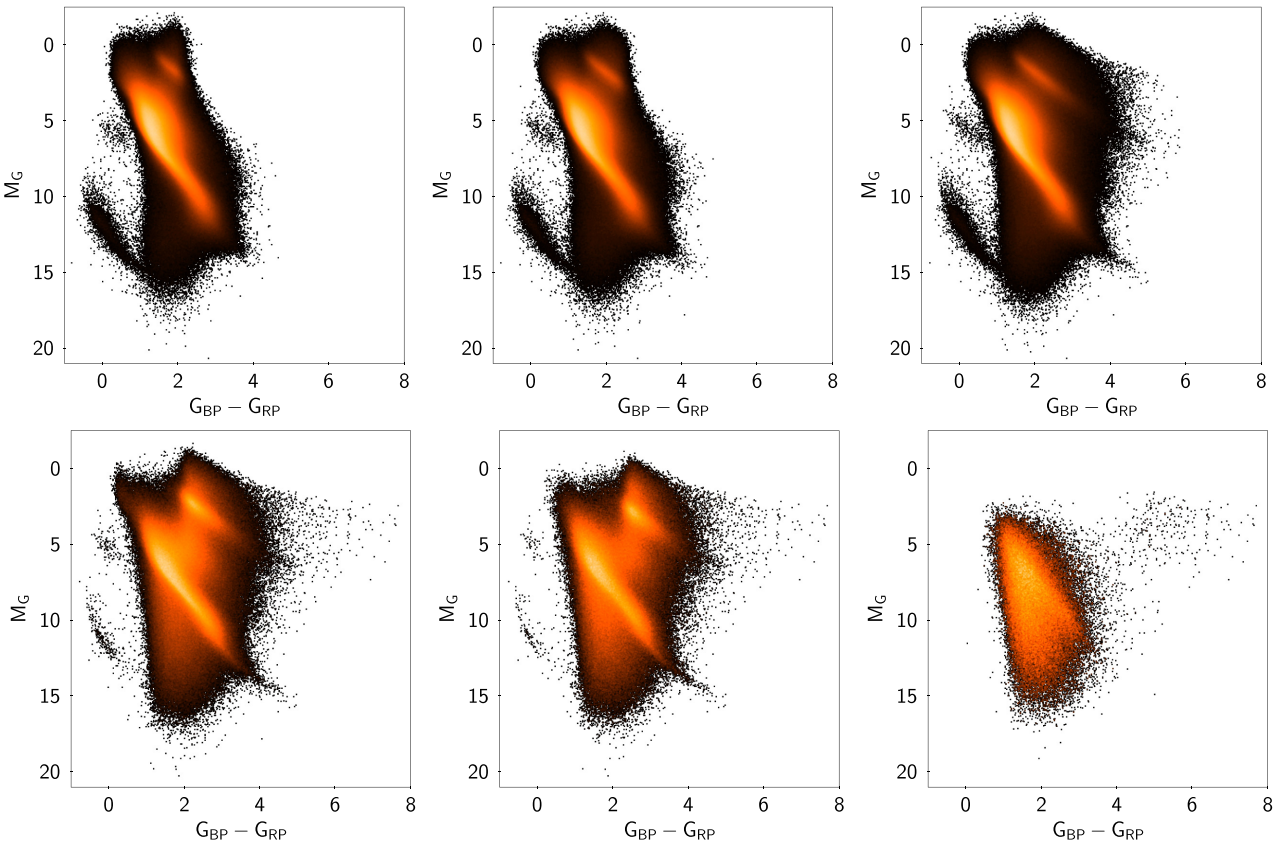

Figure 9. *Gaia* CMDs of our *Gaia*/IPHAS catalogue with varying threshold cuts on our defined 'completeness' parameter (f_c). From left to right the cuts employed are $f_c < 0.8$, $f_c < 0.9$, and $f_c < 0.99$. The top panels show the retained sources adopting a specific cut, while the bottom panels show the removed sources.

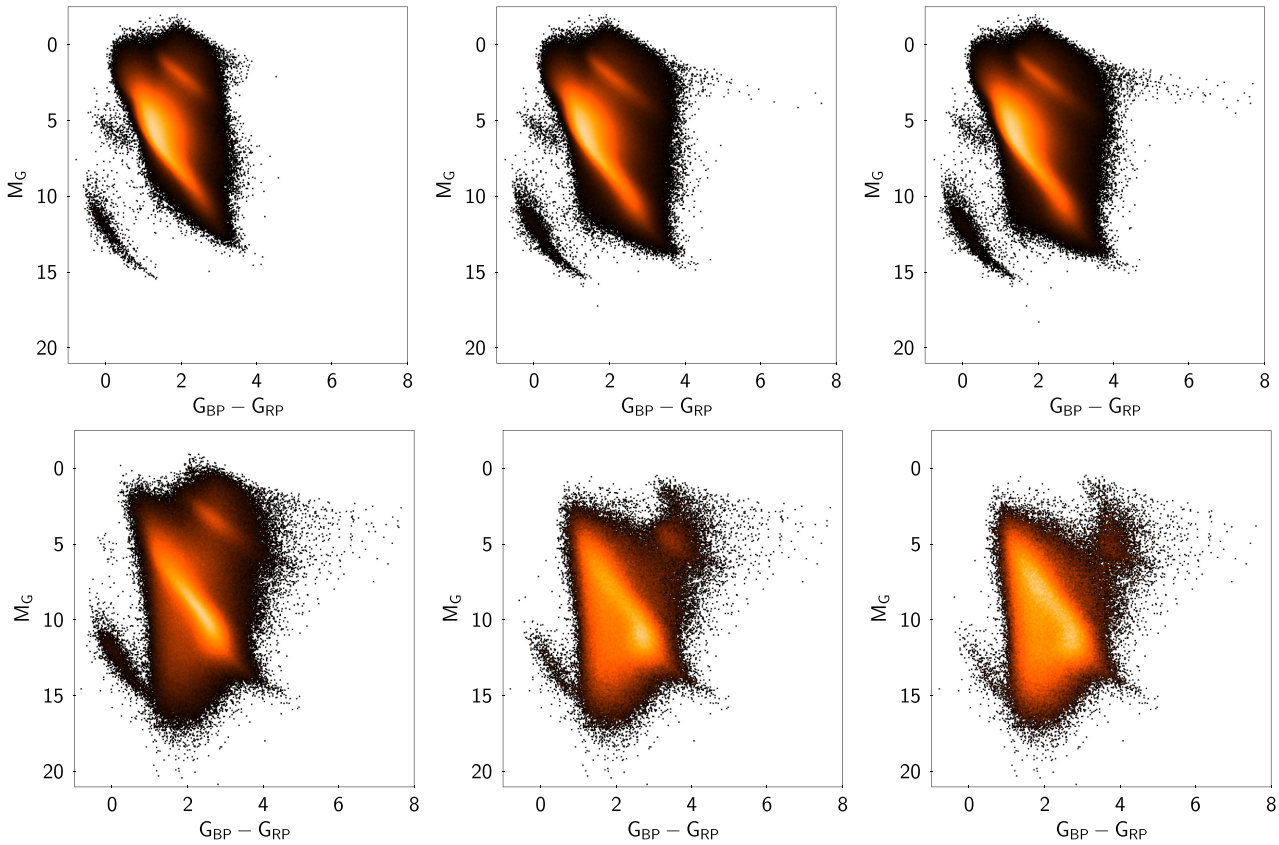

Figure 10. *Gaia* CMDs of our *Gaia*/IPHAS catalogue with varying threshold cuts on our defined false-positive parameter (f_{FP}). From left to right the cuts employed are $f_{FP} = 0$, $f_{FP} \leq 0.01$, and $f_{FP} \leq 0.02$. The top panels show the retained sources adopting a specific cut, while the bottom panels show the removed sources.

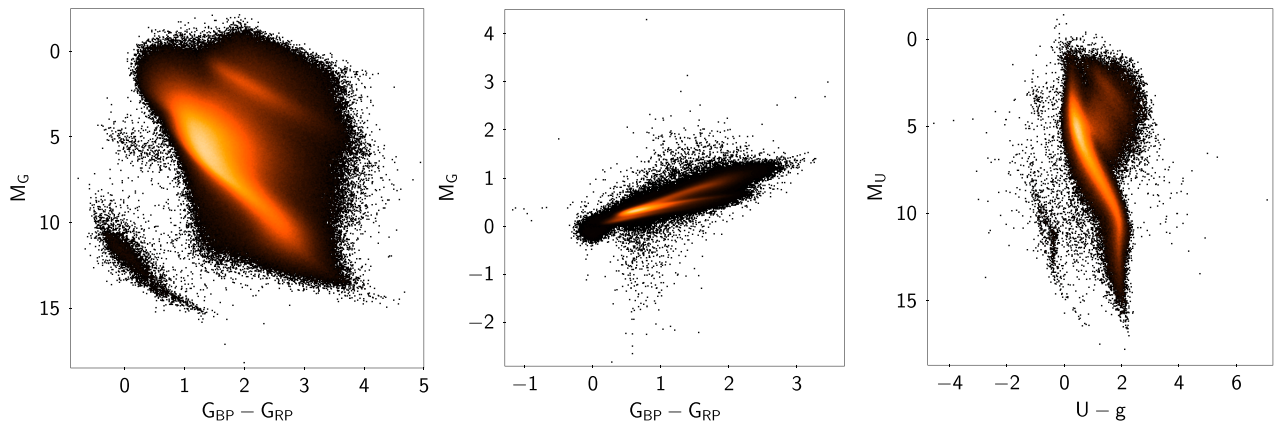

Figure 11. CMDs and colour–colour diagrams from targets retained after adopting our recommended quality cuts on f_c and f_{FP}. Left-hand panel: *Gaia*-based CMD from our *Gaia*/IPHAS catalogue. Middle panel: IPHAS colours from our *Gaia*/IPHAS catalogue. Right-hand panel: KIS-based CMD from our *Gaia*/KIS catalogue.

Fig. 12 shows the *r*, *G*, and parallax distributions for the *Gaia*/IPHAS catalogue adopting our recommended quality cuts. From these distributions we can comment that the crossmatched *Gaia*/IPHAS catalogue becomes incomplete for source fainter than $\simeq$ 16–17 mag, which expressed in terms of distance is growing incompleteness between 1 and 2 kpc. In directions of high extinction such as towards the Aquila and Cygnus Rifts, the distance turnover will be closer than elsewhere. This property is primarily a consequence of the `parallax_over_error>5` selection criterion (see Section 2.1).

4 ENABLING ADDITIONAL SCIENCE EXPLOITATION WITH *GAIA* AND IPHAS/KIS

In this section we will highlight some of the possible science exploitations possible with the value-added *Gaia*/IPHAS and

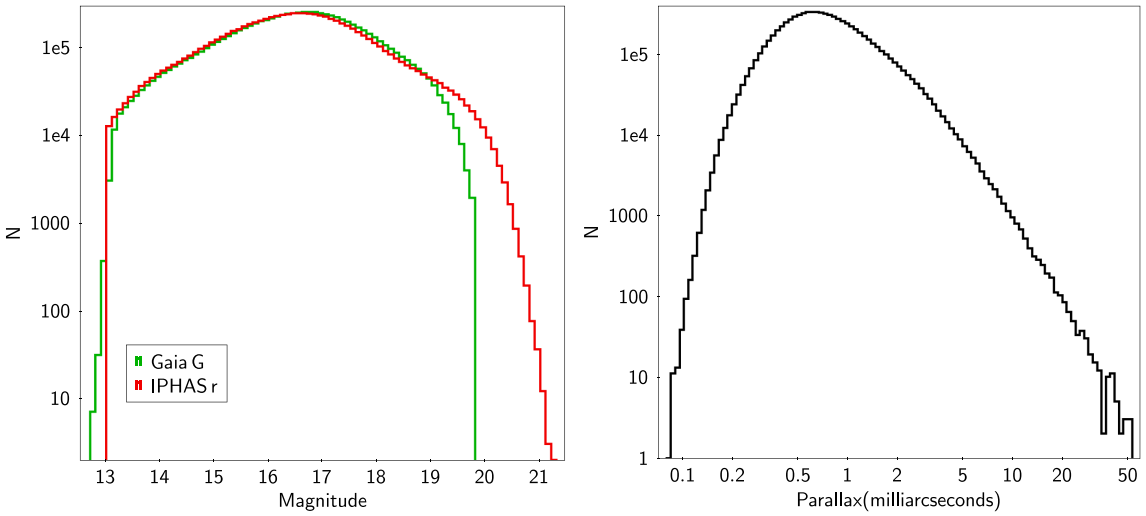

Figure 12. *r* and *G* band magnitude distributions (left-hand panel) and parallax distribution (right-hand panel) from the *Gaia*/IPHAS catalogue for all objects satisfying our quality control cuts discussed in Section 3.3.

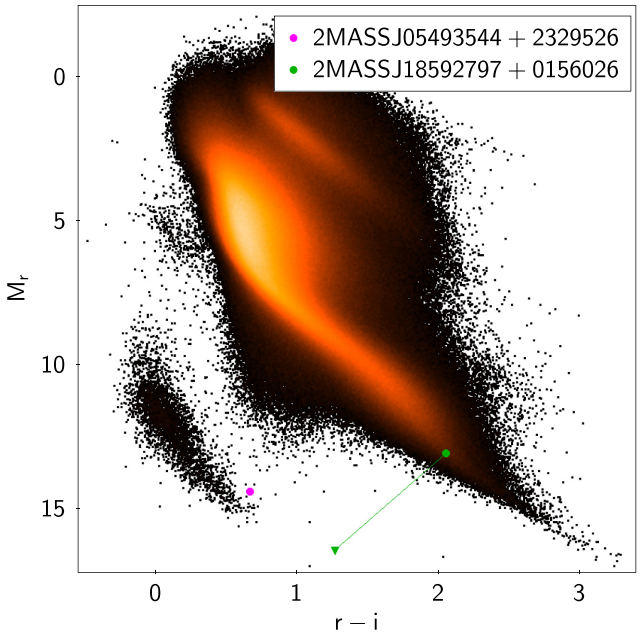

Figure 13. IPHAS-based CMD using our recommended quality cuts showing the positions of two example sources which would have either been missed (magenta point) entirely or mismatched (green point). The green line connecting the green triangle to the green point demonstrates the change in CMD position for the mismatched object in question. The IPHAS images for these targets are shown in Fig. 14.

Gaia/KIS catalogues. As in the previous section we will take examples from the *Gaia*/IPHAS catalogue, but the same exercises can also be performed with the *Gaia*/KIS catalogue.

4.1 Identifying high proper motion objects

Because of the sub-arcsecond crossmatching precision between *Gaia* and IPHAS/KIS, it is now possible to gather additional photometric information for some of the highest proper motion objects.

Fig. 13 shows our cleaned *Gaia*/IPHAS sources in the absolute IPHAS CMD diagram using the distances inferred from the *Gaia* parallaxes. Two objects are marked for illustration purposes, both also appearing in Fig. 2. One of these would have been entirely missed if proper motions were not taken into account (2MASSJ05493544+2329526 - Gaia DR2 3427482725113315200 - IPHAS DR2 J054936.18+232944.2), while the other would have been mismatched using a 5 arcsec crossmatch between *Gaia* (epoch J2000) and IPHAS (2MASSJ18592797+0156026 - Gaia DR2 4268571049773025024 - IPHAS DR2 J185928.18+015558.4). For the latter, we additionally mark the change in position within the CMD between the wrong and correct match. Fig. 14 shows, for each of the above-mentioned targets, the IPHAS *r*-band image. Also marked are the 1 arcsec crossmatch circles centred on the IPHAS recorded positions (dashed circles) and the 5 arcsec circles centred on the *Gaia* J2000.0 coordinates (solid circles).

In the case of 2MASSJ05493544+2329526, no match is found when rewinding the *Gaia* astrometry to epoch J2000 because, when the IPHAS observation was made, the target had already left the 5 arcsec radius. However, taking the IPHAS epoch of observation into account, together with the *Gaia* proper motions, reveals 2MASSJ05493544+2329526 to lie on the cold end of the white dwarf track at a distance of just under 40 pc and moving with a transverse velocity of just under 100 km s^{-1}.

In another case, 2MASSJ05210188+3425116 had also left the 5 arcsec *Gaia* target radius at IPHAS epoch of observation. However, because another faint source happened to lie within that same radius, a different IPHAS source had been associated to the *Gaia* target. Correcting for the IPHAS observation epoch reveals 2MASSJ18592797+0156026 to lie on the M-dwarf end of the main sequence, at a distance of ≈25 pc, and to be moving with a transverse velocity of ≈80 km s^{-1}.

The VACs contain many more correctly recovered targets which can now be inspected with high confidence similar to the above-mentioned examples.

4.2 Identifying 'hidden' H α excess sources

One of the biggest strengths of IPHAS (as well as KIS) is the inclusion of the narrow H α photometric band. This has two uses: (i) the $(r - H\alpha)$ colour, combined with a broad-band colour such as $(r - i)$ or $(g - i)$, enables a fix on intrinsic colour and extinction for

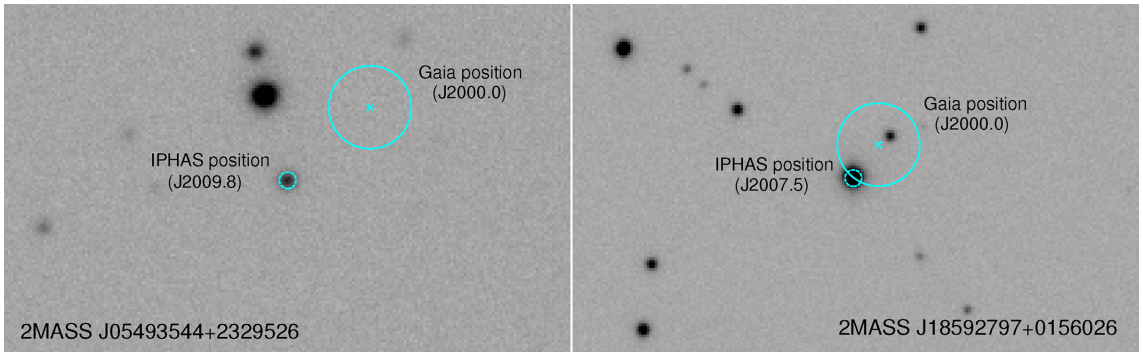

Figure 14. IPHAS *r*-band image of the recovered target 2MASSJ05493544+2329526 (left-hand panel) and 2MASSJ18592797+0156026 (right-hand panel). The dashed 1 arcsec circles are centred on the recorded IPHAS positions. The solid 5 arcsec circles are centred on the rewinded J2000.0 *Gaia* positions of the same target. The positions in the IPHAS CMD for these targets are shown in Fig. 13.

the majority of detected sources (see the discussion of this and its exploitation for extinction mapping initiated by Sale et al. 2009), (ii) large numbers of emission line stars are made evident when the $(r - H\alpha)$ colour is strong enough to represent an excess relative to the main stellar locus – removing the need for large-scale spectroscopic surveys (see e.g. Witham et al. 2006, 2007, 2008; Raddi et al. 2013; Scaringi et al. 2013; Gkouvelis et al. 2016). However, prior to the release of *Gaia* DR2, the selection of Hα-excess source candidates has usually been relative to the population of targets within a specific patch of sky, making no distinction between different luminosity classes. Most previously selected Hα-excess sources have been identified because their Hα excess places them above the unreddened main-sequence on the IPHAS $r - i$ versus $r -$Hα colour–colour diagram. An obvious drawback is that distant Hα-excess systems – especially those behind large extinction columns – will not stand out at moderate or small Hα-excess, since such objects will have the same colours as less reddened, later-type main-sequence stars in the (r-i) versus (r-Hα) colour plane (see Fig. 15, right-hand panel).

The additional parallax (and thus distance) information provided by *Gaia* DR2 now provides luminosity information that allows us to select Hα-excess sources relative to specific regions in CMD space. To illustrate the potential of this method, we have manually selected objects lying on the red-clump reddening strip from our cleaned VAC in Fig. 15, left-hand panel. The corresponding IPHAS colours are plotted in the right-hand panel of this figure.

A number of sources within this subset of systems clearly exhibit Hα-excess. Although a proportion of these Hα-excess sources would have been picked out previously as they lie above the unreddened main sequence, many others are 'hidden' below it. Using the *Gaia* distance information, and selecting Hα-excess sources based on a selection of higher-luminosity targets from the *Gaia* CMD, allows the identification of these systems as Hα-excess sources.

We illustrate this using three examples marked in Fig. 15 as solid circles along the reddened red-clump. These three systems have been selected since they are bright and have reliable LAMOST spectra (Luo et al. 2016), with a signal-to-noise greater that 10 in the r-band wavelength range and IPHAS photometry fainter than 13 in all bands (which ensures we are fainter than the non-linearity close to saturation). One of these is a clear Hα emitter as it clearly lies above the unreddened main sequence track. The other two targets would have been difficult to identify as Hα-excess sources without the additional parallax/distance information since they fall below the unreddened main sequence track in the IPHAS colour plane.

The LAMOST spectra for these three objects are shown in Fig. 16, and all present clear Hα emission lines. Based on their spectra we classify both LAMOST J045814.95+414209.4 and LAMOST J053322.50+310250.2 as Be or Herbig stars with discs. LAMOST J043749.67+514255.8 is most probably an Ae star, given the combination of Hα emission and strong Calcium triplet absorption in its spectrum. All three are rare objects in the catalogue in that they are early-type objects with distances in excess of $\sim$2 kpc (based on their parallaxes), and visual extinctions, $A_V > 3$ (based on their IPHAS $(r - i)$ colours, assuming intrinsic colours of $\sim$0 or less).

4.3 Identifying new variable sources

Our *Gaia*/IPHAS and *Gaia*/KIS VACs can also be used to identify new variable stars.

As an illustration, we have inspected the $G_{RP} - i$ colour distribution of sources in our catalogues and selected targets with $|G_{RP} - i| > 1$. This provides a useful proxy for large-amplitude variability, since the IPHAS *i*-band filter lies entirely within the *Gaia* G_{RP}-band one. There are 104 sources which satisfy both this variability criteria and the quality control cuts discussed in Section 3.3.

As discussed in Arenou et al. (2018) and Riello et al. (2018), the *Gaia* G_{BP} and G_{RP} fluxes are obtained from simple integration of a 3.5 by 2.1 arcsec window, and it is possible that close sources contaminate this measurement. The *Gaia* team thus also provide a so-called `phot_bp_rp_excess_factor` value for each source, which tries to measure the excess flux of the G_{BP} and G_{RP} bands when compared to the broader G band. Only one target out of the 104 sources found to be variable has a `phot_bp_rp_excess_factor`>2, with a sample mean of 1.45. For comparison, a mean value of 1.35 is found for our whole catalogue satisfying our quality control cuts. Visual inspection of the fields for many of these candidate variable targets reveal that these are mostly isolated sources, but there are a few exceptions. These targets will have to be followed up to be confirmed as true variables.

5 THE *GAIA*/IPHAS AND *GAIA*/KIS VALUE-ADDED CATALOGUES

We provide our full VACs without any quality selection cuts so that users can devise their own preferred selection based on f_c and f_{FP} if they so wish. These include 7927 224 and 791 071 entries for *Gaia*/IPHAS *Gaia*/and KIS, respectively.

In addition, we provide a 'light version' of the VACs, with a reduced number of columns and our recommended general-

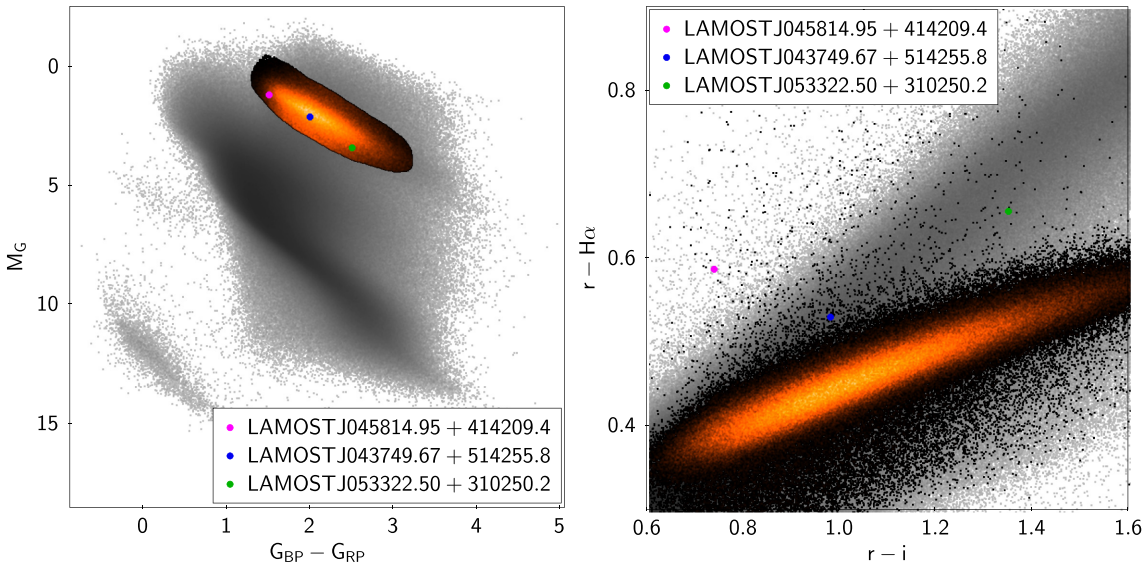

Figure 15. *Gaia*-based CMD and IPHAS colour–colour diagram for targets in our *Gaia*/IPHAS catalogue after applying our recommended quality cuts (grey points). Highlighted in red/black is the position of the reddened red-clump track within the *Gaia* CMD and its corresponding location in the IPHAS colour–colour diagram. Additionally marked are the locations of three objects which we discuss in the text, and for which LAMOST spectra exist (see Fig. 16).

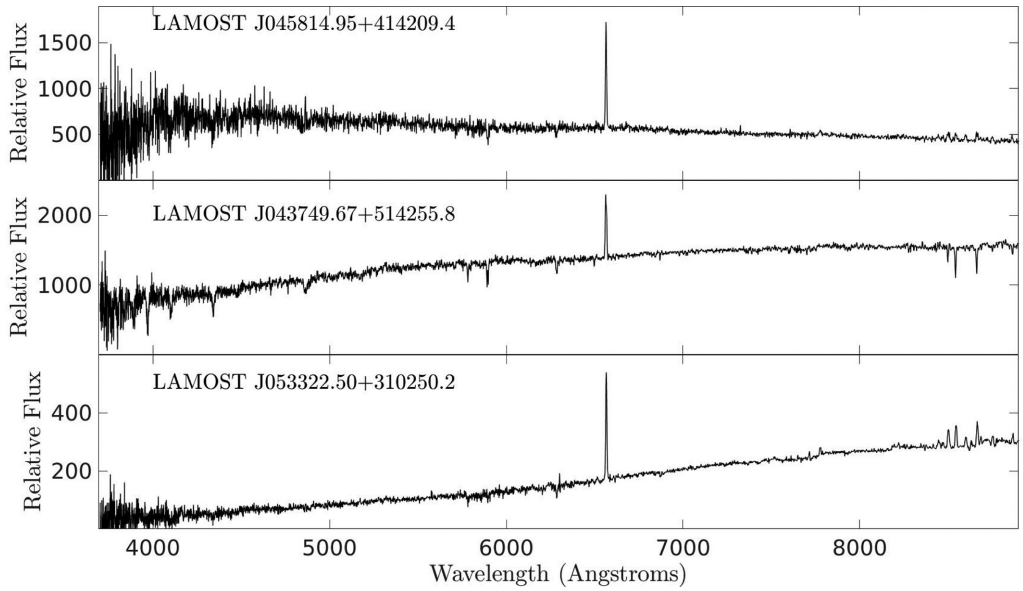

Figure 16. LAMOST spectra for the three objects picked out in Fig. 15.

purpose quality cuts already applied (see Section 3.3). These include 7479 991 (≈94 per cent) and 773 464 (≈98 per cent) entries for *Gaia*/IPHAS *Gaia*/and KIS, respectively.

The column definitions for our VAC are provided in Appendix A. The same table notes whether the columns can be found in the *Gaia*/IPHAS or *Gaia*/KIS catalogues, as well as whether the column is included in our value-added light versions of the catalogues. All catalogues can be accessed online from the VizieR service.

6 CONCLUSION

We have presented a sub-arcsecond crossmatch between *Gaia* DR2 and both IPHAS and KIS. This was achieved by taking into account both the proper motions reported in *Gaia* as well as the observation epochs reported in both IPHAS and KIS. Our VACs also contain two additional quality control parameters: a so-called completeness fraction (f_c) which provides information relating to how acceptable the *Gaia* astrometric solution is compared to targets with similar *G*-band magnitudes, and a so-called false-positive fraction (f_{FP}) providing information on how reliable the astrometric measurements of a given target are. We provide both the full catalogues with all entries, as well as a light version containing targets which satisfy our preferred quality control cuts. Aside from providing the additional IPHAS and KIS photometry to the *Gaia* information for individual targets, our catalogues can be used identify and select Hα-excess emission line sources in any part of the *Gaia* CMD as well as new variable targets.

The full *Gaia*/IPHAS catalogue contains 7927 224 targets of which 7479 991 pass our quality control cuts, while the *Gaia*/KIS catalogue contains 791 071 targets of which 773 464 pass the quality cuts. Both the full and light versions of the catalogue can be obtained through VizieR.

ACKNOWLEDGEMENTS

Crossmatching between catalogues has been performed using STILTS, and diagrams were produced using the astronomy-oriented data handling and visualization software TOPCAT (Taylor 2005). This work has used the Astronomy & Astrophysics package for Matlab (Ofek 2014). This research has also made extensive use of the SIMBAD data base, and our catalogue is being hosted on VizieR, both operated by the Centre de Donnees astronomiques de Strasbourg (CDS, Ochsenbin, Bauer & Marcout 2000). JED and MM acknowledge the support of a research grant funded by the Science, Technology and Facilities Council of the UK (STFC, ref. ST/M001008/1).

REFERENCES

Arenou F. et al., 2018, A&A, 616, A17
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, AJ, 156, 58
Barentsen G. et al., 2014, MNRAS, 444, 3230
Drew J. E. et al., 2005, MNRAS, 362, 753
Farnhill H. J., Drew J. E., Barentsen G., González-Solares E. A., 2016, MNRAS, 457, 642
Gaia Collaboration, 2016, A&A, 595, A1
Gaia Collaboration, 2018a, A&A, 616, A1
Gaia Collaboration, 2018b, A&A, 616, A10
Gkouvelis L., Fabregat J., Zorec J., Steeghs D., Drew J. E., Raddi R., Wright N. J., Drake J. J., 2016, A&A, 591, A140
González-Solares E. A. et al., 2008, MNRAS, 388, 89
Greiss S. et al., 2012, AJ, 144, 24
Groot P. J. et al., 2009, MNRAS, 399, 323
Lindegren L. et al., 2018, A&A, 616, A2
Luo A. L. et al., 2016, VizieR Online Data Catalog p. V/149
Luri X. et al., 2018, A&A, 616, A9
Ochsenbin F., Bauer P., Marcout J., 2000, A&AS, 143, 23
Ofek E. O., 2014, Astrophysics Source Code Library, record ascl:1407.005
Raddi R. et al., 2013, MNRAS, 430, 2169
Riello M. et al., 2018, A&A, 616, A3
Sale S. E. et al., 2009, MNRAS, 392, 497
Scaringi S., Groot P. J., Verbeek K., Greiss S., Knigge C., Körding E., 2013, MNRAS, 428, 2207
Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, ASP Conf. Ser. Vol. 347. Astronomical Data Analysis Software and Systems XIV. Astron. Soc. Pac, San Francisco, p. 29
Witham A. R. et al., 2006, MNRAS, 369, 581
Witham A. R. et al., 2007, MNRAS, 382, 1158
Witham A. R., Knigge C., Drew J. E., Greimel R., Steeghs D., Gänsicke B. T., Groot P. J., Mampaso A., 2008, MNRAS, 384, 1277

SUPPORTING INFORMATION

Supplementary data are available at http://cdsarc.u-strasbg.fr/viz-bin/qcat?I/36.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: CATALOGUE FORMAT

Table A1. Definition of columns in the *Gaia*/IPHAS and *Gaia*/KIS catalogues. Column indices are given for both the light and full catalogue versions. Light versions only contain a limited number of columns and are restricted to sources satisfying our recommended quality cuts of $f_c < 0.98$ and $f_{FP} \leq 0.02$. Absolute magnitudes and transverse velocities have been computed using the 1/parallax method. It is important to point out that the absolute magnitude and colour columns in the catalogue do not take extinction/reddening into account.

Gaia/IPHAS (full)	*Gaia*/IPHAS (light)	*Gaia*/KIS (full)	*Gaia*/KIS (light)	Column name	Unit	Description
1	1	1	1	GaiaDR2		Unique *Gaia* DR2 source designation
2	2	2	2	ra	degrees	*Gaia* DR2 barycentric right ascension (ICRS) at Epoch 2015.5
3	3	3	3	dec	degrees	*Gaia* DR2 barycentric declination (ICRS) at Epoch 2015.5
4	–	4	–	e_ra	mas	Standard error of right ascension (raErr×cos(dec))
5	–	5	–	e_dec	mas	Standard error of declination
6	4	6	4	Plx	mas	Absolute stellar parallax
7	–	7	–	e_Plx	mas	Standard error of parallax
8	5	8	5	pmra	mas yr^{-1}	Proper motion in right ascension direction (raPM×cos(dec))
9	6	9	6	pmdec	mas yr^{-1}	Proper motion in declination direction
10	–	10	–	e_pmra	mas yr^{-1}	Standard error of proper motion in right ascension direction
11	–	11	–	e_pmdec	mas yr^{-1}	Standard error of proper motion in declination direction
12	7	12	7	G	mag	*Gaia* *G*-band magnitude
13	8	13	8	BP	mag	*Gaia* G_{BP}-band magnitude
14	9	14	9	RP	mag	*Gaia* G_{RP}-band magnitude
–	–	15	10	U	mag	KIS *U*-band magnitude
–	–	16	–	e_U	mag	Standard error on KIS *U*-band magnitude
–	–	17	11	g	mag	KIS g'-band magnitude
–	–	18	–	e_g	mag	Standard error on KIS g'-band magnitude
15	10	19	12	r	mag	IPHAS or KIS r'-band magnitude
16	–	20	–	e_r	mag	Standard error on IPHAS or KIS r'-band magnitude
17	11	21	13	i	mag	IPHAS or KIS i'-band magnitude
18	–	22	–	e_i	mag	Standard error on IPHAS or KIS i'-band magnitude
19	12	23	14	ha	mag	IPHAS or KIS Hα-band magnitude
20	–	24	–	e_ha	mag	Standard error on IPHAS or KIS Hα-band magnitude
21	–	25	–	BPmRP	mag	$G_{BP} - G_{RP}$ colour
22	–	26	–	rmi	mag	$(r' - i')$ colour
23	–	27	–	rmha	mag	$(r' -$H$\alpha)$ colour
–	–	28	–	Umg	mag	$(U - g')$ colour
24	–	29	–	pmT	mas yr^{-1}	Transverse proper motion
25	–	30	–	vT	km s^{-1}	Transverse velocity
26	13	31	15	mMJD	d	Modified Julian Date used for crossmatching *Gaia* to IPHAS and KIS.
27	–	32	–	mMJD_separation	arcsec	Angular separation between the rewinded *Gaia* position at Epoch mMJD to the nominal IPHAS or KIS position
28	–	33	–	M_G	mag	Absolute *Gaia* *G*-band magnitude (not corrected for extinction)
29	–	34	–	M_r	mag	Absolute IPHAS r'-band magnitude (not corrected for extinction)
–	–	35	–	M_U	mag	Absolute KIS *U*-band magnitude (not corrected for extinction)
30	–	36	–	redChi2		Reduced χ^2 (χ^2_ν) obtained from the *Gaia* astrometric fit
31	14	37	16	f_c		Retention fraction based on χ^2_ν (completeness)
32	15	38	17	f_FP		False-Positive rate based on negative parallax sample
33	–	39	–	ra_mMJD	degrees	Right ascension provided by IPHAS or KIS DR2 (ICRS, Epoch mMJD)
34	–	40	–	dec_mMJD	degrees	Declination provided by IPHAS or KIS DR2 (ICRS, Epoch mMJD)
35	–	–	–	fieldID		IPHAS DR2 field identifier
36	–	–	–	IPHAS_name		IPHAS DR2 name
–	–	41	–	KIS_name		KIS DR2 name
–	–	42	–	KIC		KIC ID

This paper has been typeset from a TEX/LATEX file prepared by the author.